

AD-A103 341

RAND CORP SANTA MONICA CA
QUALITY OF CARE IN EPISODES OF COMMON RESPIRATORY INFECTIONS IN--ETC(U)
OCT 80 K N LOHR

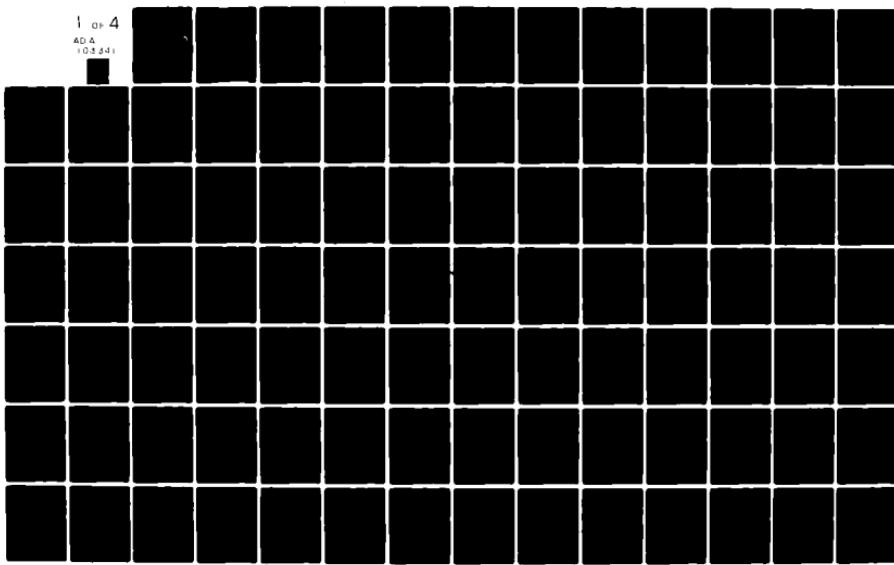
F/G 6/5

UNCLASSIFIED RAND/P-6570

NL

1 of 4

ADA
102341



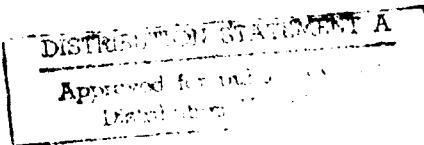
LEW ✓
1
AD A103341

QUALITY OF CARE IN EPISODES OF COMMON RESPIRATORY
INFECTIONS IN A DISADVANTAGED POPULATION

KATHLEEN NIES LOHR



111 OCTOBER 1980



81 8 25 104 P-6570
52
1166

The Rand Paper Series

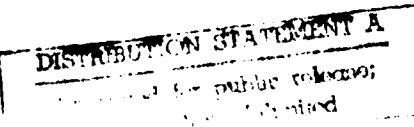
Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

QUALITY OF CARE IN EPISODES OF COMMON RESPIRATORY
INFECTIONS IN A DISADVANTAGED POPULATION

KATHLEEN NIES LOHR

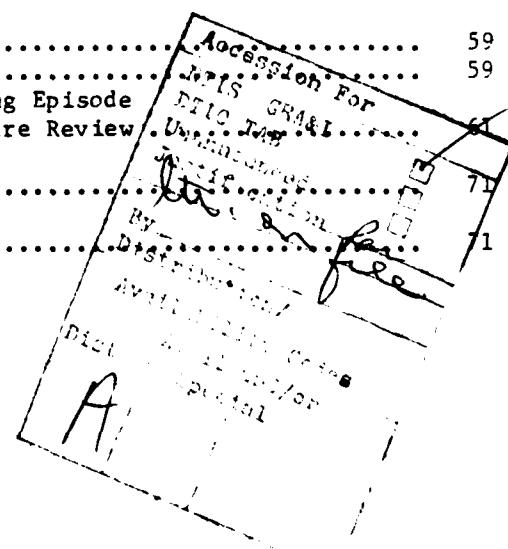
OCTOBER 1980



CONTENTS

ACKNOWLEDGMENTS	ix
CHAPTER I. INTRODUCTION AND BACKGROUND	1
Peer Review and Quality Assessment:	
Overview of the Literature	3
Peer Review and the New Mexico EMCRO	4
General Methods Issues	7
Validity and Generalizability	7
Time Periods and Cohort Analyses	10
Previous Evaluations of the Ambulatory Care Review Activities	12
First (Two-Year) Study: All Injectable Drugs	12
Second (Four-Year) Study: All Injectable Drugs	14
Third (Two-Year) Study: Injectable Antibiotics	16
Third (Two-Year) Study: Episodes of Care	18
CHAPTER II. VALIDATION OF EPISODE RULES	29
Rand Health Insurance Study Data	30
Methods of Abstracting HIS Episodes	33
Splitting the HIS Episode Sample	34
Issues in Abstracting HIS Sore Throat Episodes	35
Results of Abstracting HIS Sore Throat Episodes	39
Validation Studies	47
Methods of Applying the New Mexico Rules to HIS Episodes	47
Results of Applying the New Mexico Rules to HIS Episodes	48
Discussion	59
Episode Algorithms	59
Restricting the Data Set	59
Implications of Choosing Among Episode Diagnoses for Quality-of-Care Review	60
CHAPTER III. QUALITY-OF-CARE CRITERIA	71
Introduction	71

PRECEDING PAGE BLANK-NOT FILLED



General Considerations	72
Terminology	74
Methods	75
Physician Judges	75
The Questionnaire	75
Median Ratings	79
Agreement Among Judges	79
Diagnosis-Specific Profiles of Quality of Care	81
Results	83
Judges and Questionnaires	83
Median Ratings	85
Agreement Among Judges	93
Quality-of-Care Profiles	97
Discussion	106
CHAPTER IV. QUALITY OF CARE IN EPISODES OF RESPIRATORY INFECTION ..	111
Methods	113
Creating Episodes of Care	113
Scoring by Profiles of Care	119
Quality-of-Care Analyses	121
Results	121
Distribution of Episodes	121
A Note Regarding Levels of Quality-of-Care	123
Differences in Quality of Care Among Types of Providers in Periods I and II: Five Respiratory Infections	124
Differences in Quality of Care Between Periods I and II	136
Comparisons Between the Earlier and Present Studies	140
Management of Sore Throat Episodes	141
Differences in Quality of Care Among Types of Providers in Periods I and II: Otitis Media	145
Summary	153
Discussion	154
Factors Leading to Good and Bad Care	154
A "Quasi-Continuum" of Care	155
Improved Care Versus Lingering Inadequacies	157

CHAPTER V. CONCLUSIONS AND POLICY IMPLICATIONS	159
Principal Findings	161
Determining Episodes of Care	161
Diagnosis-Specific Profiles For Judging Quality of Care ..	163
Quality-of-Care Analyses	164
Discussion	167
Episodes of Care	168
Profiles of Care	170
Future Directions of PSRO Ambulatory Care Review	174
Summary	182
APPENDICES	185
APPENDIX A. VALIDATION OF EPISODE RULES	186
Rand Health Insurance Study Data	186
Abstracting HIS Episodes	191
APPENDIX B. METHODOLOGIC DETAILS OF DESIGNING THE QUALITY-OF-CARE PROFILES	194
Methods	194
Questionnaires	194
Median Ratings	195
Agreement Among Judges	198
APPENDIX C. EPISODE RULES	236
Produce Final Workfile	236
Create Episodes	240
APPENDIX D. DIAGNOSIS-SPECIFIC DISTRIBUTIONS OF EPISODES BY LEVEL OF QUALITY OF CARE FOR SEPARATE CATEGORIES OF PROVIDERS IN PERIODS I AND II	249
APPENDIX E. USE OF THROAT CULTURES AND APPROPRIATE ANTIBIOTICS IN SORE THROAT EPISODES FOR SEPARATE CATEGORIES OF PROVIDERS IN PERIODS I AND II	282
REFERENCES	285

ACKNOWLEDGMENTS

For their unfailing support and thoughtful guidance in this work, I am greatly indebted to Joseph Newhouse, John Rolph, and, especially, Robert Brook, who was always willing to answer one more question or read one more draft chapter. I do not exaggerate in saying that, without their assistance, this work would not have come to fruition.

I wish to acknowledge the physicians who contributed their time and expertise to the quality-of-care ratings: Robert Bourne, Robert Brook, George Goldberg, Thomas Inui, Robert Kane, Michael Kaufman, John King, James LoGerfo, Samuel Sapin, Bonnie Scott, and Stephen Williams.

Several Rand colleagues and friends were helpful at one time or another during my sojourn through The Rand Graduate Institute, among them Alvin Harman, Carl Morris, and John Ware, as well as Allyson Davies, Caren Kamberg, and Meinaard Veen, and I thank them for their encouragement and good advice. I am grateful also to Charles Wolf, Jr., and Mary Anderson for making financial aid available to complete parts of this study.

Throughout this work, Barbara Eubank cheerfully prepared more drafts of tables and text than any one person should ever be expected to do, especially under great time pressure, and I am grateful for her skills and her equanimity. Paul Norton did a commendable job of programming.

Finally, I am deeply grateful to my family for their unlimited love, patience, and help during what seemed, at times, to be an endless task. My husband, Bill, was unstintingly supportive and determined to see it through with me. My children, Chris and Michael, and later Sam, Lee, and Kirsten, showed understanding beyond their years. My parents, Kathleen and Elmer Nies, never faltered in providing encouragement. They all know, I believe, how much I thank them.

The original version of this study was prepared as a dissertation in partial fulfillment of the requirements of the doctoral degree in policy analysis at The Rand Graduate Institute.

The faculty committee that supervised and approved the dissertation consisted of Robert H. Brook, Chairman, Joseph P. Newhouse, and John E. Rolph.

CHAPTER I
INTRODUCTION AND BACKGROUND

This monograph examines the quality of care delivered to New Mexico Medicaid patients with common respiratory infections and the effect of the New Mexico Experimental Medical Care Review Organization (EMCRO) on that care in the early 1970s. EMCROs were precursors of Professional Standards Review Organizations (PSROs). PSROs are physician-led peer review organizations for federally funded medical care programs such as Medicaid and Medicare.

For these analyses, quality of care is viewed comprehensively; criteria are defined as "profiles" that simultaneously include physician visits, laboratory tests and other procedures, and medications in injectable and oral form. Care is evaluated for entire episodes of care developed from computerized Medicaid claims data.

These analyses proceed from studies carried out over the past four years; each study has built on previous work. To give some perspective to the latest work, this chapter briefly reviews earlier evaluations and research activities using Medicaid data from the New Mexico EMCRO. Summary findings on quality of care cover use of injectable drugs generally, use of injectable antibiotics by different types of physicians, and use of both injectable and oral antibiotics for specific diagnoses by different physician types. In addition, much of the present work uses a refined set of computerized rules for assigning services to entire episodes of care for respiratory infections; the

original set of rules is described to provide a context in which the present (modified) rules can be understood.

Beyond this chapter, the report is organized as follows: Chapter II presents the methods and results of efforts to validate and where necessary refine the computer rules for creating episodes of care. These validation studies were carried out using external, more complete data--namely, two years of ambulatory care data from the insurance claim files of the Rand Health Insurance Study (HIS). The HIS is a long term social experiment to provide information on how differences in the price of medical care influences use of services, health status, and quality of care.

Chapter III presents the methods and results of establishing the quality-of-care "profiles" for the six respiratory infections of interest. A questionnaire was fielded among selected physicians in private and academic practice to elicit ratings as to the acceptability or nonacceptability of various medications, procedures, and other services for these particular conditions. Then from the physician panel's ratings, profiles were developed that simultaneously took into account the presence and absence of acceptable and unacceptable services.

Chapter IV describes the methods for creating episodes from the Medicaid claims data and for evaluating quality of care by using the diagnosis-specific profiles. It documents differences in quality of care (1) as a function of certain physician characteristics and (2) before and after the New Mexico EMCRO had begun quality-related peer review activities; also reported are special analyses to clarify

elements of high and poor quality of care. Chapter V provides a brief synopsis of the present work and a discussion of its implications for quality-of-care assessment and the role of PSROs in ambulatory care review.

PEER REVIEW AND QUALITY ASSESSMENT:

OVERVIEW OF THE LITERATURE

Concern for the quality of medical care has a very long history, reaching back in modern times at least to Florence Nightingale; the quality assessment and assurance movement as we know it today was originated by researchers such as Sheps, Donabedian, Clute, and Peterson over two decades ago. Since that time, the literature in the field has become voluminous. Principal issues in the field include documenting deficiencies in quality of both inpatient and ambulatory care, targeting quality-of-care activities, defining and measuring quality, implementing quality assurance systems, and estimating the presumed costs and benefits of quality-related efforts.

This literature is much too vast to review here; the citations below are to selected bibliographies and reference works that adequately cover the major works in the field. The history of quality assessment techniques can be found, for example, in Brook (1973), Lewis (1974), Christoffel and Loewenthal (1977), Brook et al. (1977), or McAuliffe (1979), and that of quality assurance organizations such as Foundations for Medical Care, EMCROs, and PSROs in Toman et al., (1976), Brook and Williams (1976), Brook and Davies-Avery (1977), or Mushlin and Appel

(1980). Works oriented more toward policy issues surrounding utilization review, peer review, and quality assurance include those of the Institute of Medicine (IOM, 1974, 1976), various federal health agencies (Goran et al., 1975; Goran, 1979; OPEL, 1977; HCFA, 1980), and other agencies or researchers (Greene, 1976; Kessner, 1978; Gertman et al., 1979; CBO, 1979).

Williamson (1977) compiled and annotated a comprehensive bibliography of the existing literature to the middle of the 1970s. Williams and Brook (1978) and Sanazaro (1980) have produced more recent detailed state-of-the-art reviews of the literature on quality assessment and assurance techniques and relevant policy issues. By the end of the decade, numerous books and monographs also had appeared that dealt with special aspects of protecting and improving the quality of medical care (Egdahl and Gertman, 1976; 1978; Avery et al., 1977; Williamson, 1978; Hulka et al., 1979; Miller and Knapp, 1979; Riedel and Riedel, 1979).

PEER REVIEW AND THE NEW MEXICO EMCRO

The focus of this study is on the New Mexico EMCRO, which can be considered a prototype for the national PSRO program (although it was not so designed). The details of how the EMCRO functioned, including its techniques for reviewing ambulatory care claims submitted to the Medicaid program, have been documented elsewhere (Brook and Williams, 1976; Brook et al., 1978; Lohr et al., 1980). Briefly, the EMCRO was established in September 1971 and functioned through 1975 as the formal

peer review system for the statewide Medicaid program. It was formed as a three part organization--a fiscal intermediary component (the Dikewood Corporation), a peer review component (the New Mexico Foundation for Medical Care), and the State Department of Health and Social Services. The EMCRO was intended both to improve the quality of care and to control the use and cost of medical care for Medicaid eligibles in that state.

From the outset the EMCRO undertook review of both inpatient and ambulatory care by means of a comprehensive, computerized system involving both clerical and professional (physician) reviewers. To this end, the EMCRO developed a system by which Medicaid claims were reviewed against both administrative guidelines (regarding factors such as patient eligibility and duplicate claims) and quality-of-care guidelines. To review the quality of ambulatory care, clerical employees screened all ambulatory claims submitted by physicians for payment by the Medicaid program; for this screening they used guidelines that had been developed by physician members of the EMCRO. Depending on the findings of the clerical employees, the claim was either reimbursed (in part or in full) or was referred to physician reviewers for a decision as to whether further information should be requested from the physician rendering the care or payment denied for the service because it was medically inappropriate. In the event a claim was denied (i.e., payment for the service was denied for medical reasons), the reviewing physician was required to state a reason and the physician rendering the care was notified of the denial. The latter had the option of appealing the decision of the reviewer physician to an appeals review committee.

Regarding ambulatory care, the EMCRO focused on the type and frequency of injectable drugs and gave less emphasis to the type and frequency of laboratory tests, frequency and level of office visits, and the need for emergency room services. Claims for prescription drugs were not subjected to quality-of-care-related peer review.

Early in its operations, the EMCRO developed guidelines for the use of injectable drugs that were related specifically to diagnoses. In January 1972, four months into its operations, the EMCRO started educational efforts among physicians who were overusing injections. These efforts consisted mostly of informal contacts between Foundation physicians and practicing physicians who seemed to be giving many unnecessary injections. The guidelines were widely circulated at this time as well. In May 1972, the formal policy regarding the use of injections was adopted by the Foundation and disseminated in written form to all physicians participating in the Medicaid program in New Mexico. Thereafter, denial of payment for those injections given for medical reasons that did not meet the guidelines and for injections that peer review physicians felt were medically unnecessary was begun in earnest. (The injection guidelines can be found in Brook and Williams (1976) or Lohr et al. (1980).)

In keeping with its efforts to address critical medical care system problems of the decade, the EMCRO concentrated its major efforts during these years on instituting a fairly complex utilization review (UR) program for controlling inpatient hospital use. The UR system involved a pre-admission certification program for elective admissions, a certification program for emergency admissions, and a recertification

program for extended stay admissions. These activities were performed in part by hospitals that had been delegated some review authority and in part by the New Mexico Foundation for Medical Care. An evaluation of that effort (Brook et al., 1978) was undertaken partially because the type of program developed in New Mexico was similar to the program now being used by the New Mexico PSRO and other PSROs. It indicated that the EMCRO's three-part UR approach to controlling hospital use and hospital expenditure had no apparent effect on length of stay, proportion of stays exceeding the Professional Activities Study (PAS) 50th percentile, or the probability of admission to a hospital. These findings, which stand in some contrast to those related to ambulatory care, are reported in some detail in the monograph just cited and will not be reviewed further.

GENERAL METHODS ISSUES

VALIDITY AND GENERALIZABILITY

Several threats to the internal validity or generalizability of these studies should be noted. One is that these findings are not drawn from a program that was implemented in a truly experimental manner; moreover, it was not possible to obtain two to three years of appropriate "before" data. To deal with these drawbacks, the controlled time series design takes advantage of several facts: (1) During its lifetime, the New Mexico EMCRO did not review some services at all (prescription drugs) or did not review all services with equal intensity or with equal regard for quality-of-care issues; (2) it began serious

review of injections only after a few months had passed. Thus, it was possible to employ the pattern of use of the unreviewed services as a partial control in evaluating the impact of the peer review organization on the use of other services and to define a "before" period that was relatively uncontaminated by possible program effects.

The analyses might have been affected by many factors, among the more serious of which would be changes in the Medicaid population. Because of the high quality of data maintained on eligibility of persons enrolled in the Medicaid population, however, identifying a large cohort that did not change (except for aging) for the entire four years of the study was possible; it provided adequate sample sizes for these analyses. Moreover, data coded from insurance claim forms may contain errors in procedure codes, diagnoses, and so forth. Dikewood, however, had instituted several procedures to guard against such problems (such as returning illegible claims to the provider for clarification), and various validity checks of their data base indicated that a good deal of confidence could be placed in the data.

A major change in physician practices on the first day of operation of the peer review system would be missed by this method. The likelihood that this occurred was virtually nil, however. In the first place, the peer review system was implemented gradually. Moreover, large changes in physician behavior generally occur in response to systematic and considerable applications of direct quality assessment/assurance efforts, not to the early or tentative steps of a new program.

The episode methodology in the earlier work was new and untested. Validity studies were undertaken, therefore, expressly to determine how well the rules for developing episodes worked and to what degree errors and complicated sets of services within episodes might be encountered. On the basis of these studies, the episode methodology was refined (as described in Chapter II) to be more sensitive to possible biases and to take better account of the natural history of respiratory infections.

Another problem arises in using physician-generated and -validated explicit process criteria to measure quality of care. Numerous studies have shown that adherence to criteria by practicing physicians, in the sense of their performing various widely accepted criteria items, is often quite low (Payne et al., 1976; Osborne and Thompson, 1975; Hare and Barnoon, 1973; Greenfield et al., 1978; Tonkin, 1979; Sullivan et al., 1980). Thus, one can expect a priori that, when quality is measured against very comprehensive explicit criteria, it may well be judged as fairly unsatisfactory.

The research reported here had to rely on explicit process criteria. No recourse to implicit judgments based on observer ratings of patient-physician encounters or on abstracts of illness episodes is possible in a study based exclusively on claims data. The studies just cited, however, often used process criteria that had many more detailed elements than do the criteria developed for this research. This fact should decrease any likelihood that unwarranted judgments (that quality of care is low) will result solely from the criteria used here.

As will be seen in Chapters III and IV, much has been done to give physicians the benefit of the doubt in establishing the explicit quality Profiles and to allow for degrees of acceptability of care. Moreover,

explicit criteria often focus on the presence of "good" things that should be done (and recorded in a medical record), not on the absence of "bad" things. This study thus attempts to go beyond the typical external criteria approach by incorporating criteria that deal with the presence and absence of both helpful and harmful elements of care.

Despite the fact that the EMCRO can be viewed as a PSRO prototype, the findings about the New Mexico EMCRO should be generalized only with caution, because of differences between the New Mexico and the national Medicaid and peer review environments. First, New Mexico is a large state geographically but has only one major metropolitan area. Second, New Mexico's Medicaid population is mostly white (of which perhaps one-third is of Hispanic origin); Indians represent the only other substantial minority. Third, New Mexico has very few physicians who graduated from foreign medical schools and relatively high numbers of osteopaths. Fourth, there is only one university medical center. Finally, the New Mexico EMCRO ambulatory care effort was generously funded; physicians, the fiscal carrier, and state bureaucrats exploited a cooperative atmosphere to implement a comprehensive peer review system relatively swiftly.

TIME PERIODS AND COHORT ANALYSES

As noted earlier, these investigations have used a quasi-experimental time series design. "Before" and "after" periods were defined for some analyses, and differences in dependent variables such as numbers of injections billed per ambulatory visit or quality-of-care scores by provider types were investigated. Period I

was (and is for the present work) defined as the six months preceding dissemination of the EMCRO injection guidelines, i.e., September 1971 through February 1972.* Period II is September 1973 through February 1974, after most of the effect of the EMCRO on reducing injection use had been achieved.

Some earlier analyses were based on the entire Medicaid population; others, including those reported in this monograph, were done on an Aid to Families with Dependent Children (AFDC) cohort that had been enrolled in the Medicaid program continuously for four years. This cohort was chosen for several reasons: (1) AFDC eligibles were likely to be relatively healthy, in contrast to the other aid categories in Medicaid, for which eligibility criteria were either the presence of a specific chronic condition or handicap, or old age. Because it was relatively healthy, diagnosis-specific analyses on this cohort would be less confounded by the presence of underlying chronic disease. (2) No Medicaid-Medicare "cross-over" problems would be encountered in this cohort; i.e., no services were covered first by Medicare, as is true for Old Age Assistance, meaning that all necessary ambulatory care data could be contained in the Medicaid files. (3) Altogether, AFDC made up the largest portion (about 70 percent) of the Medicaid enrollees in New Mexico. (4) The AFDC cohort made up about 25 percent of Medicaid eligibles and accounted for more than 40 percent of all Medicaid dollars per eligible per year during this time.

*In some early analyses, Periods I and II covered a five-month period between September and January; the six-month period was used in the later work for methodological reasons.

PREVIOUS EVALUATIONS OF THE AMBULATORY CARE REVIEW ACTIVITIESFIRST (TWO-YEAR) STUDY: ALL INJECTABLE DRUGS

The first evaluation, done after the EMCRO had operated for two years, focused on changes over time in the use of injectable drugs and the relationship between inappropriate use of injections and selected physician characteristics that are believed to be associated with differences in quality of care (Brook and Williams, 1976). During those two years, more than 95,000 injections were administered to the Medicaid population; of these, 27,640 (29 percent) were denied for medical reasons. Close to 50 percent of the injections were antibiotics; about one-quarter of these, on average, were denied for medical reasons. Of other major categories of injections, such as antinauseants and antiemetics, steroids, or hormones, between one-fifth and three-fifths were denied for medical reasons.

In the analyses of the relationship between physician characteristics and use of injections, the main dependent variables were the numbers of injections billed and denied per ambulatory visit. The major explanatory variables for physicians in solo practice were age of physician, country of medical education, specialty, board certification status, rural/urban location of practice, type of doctor (Doctor of Osteopathy [DO] or Doctor of Medicine [MD]); for one analysis, group practice versus solo practice was also investigated.

The major findings of the first evaluation included the following. DOs had a higher rate of medically inappropriate use of injections than

MDs, and physicians who were not board-certified had a higher rate than those who were. Findings on physician age were equivocal. Quality deteriorated with age for DOs but not MDs; for the latter, the poorest care was given by the youngest and oldest physicians. Foreign medical graduates did not give significantly poorer care than U.S. graduates. Quality of care also did not vary systematically with practice location. Results with regard to specialty were mixed: the poorest care was given by obstetricians/gynecologists and the best care by pediatricians; general practitioners were not markedly different from internists. Finally, partnerships, groups, and clinics tended to provide care superior to that provided by any class of solo physician.

In this study, a small set of "outlier" physicians emerged, characterized by extremely high rates of denial of payments for injections for medical reasons. They numbered 15 MDs and 7 DOs (7 percent of the 313 solo physicians studied) and accounted for more than 40 percent of all unnecessary injections but only 14 percent of all ambulatory visits. This same set of 22 physicians has been labeled "outlier" in all subsequent studies.

As noted, the chief purpose of the first study was to evaluate EMCRO performance, by determining whether it was able to decrease the use of inappropriate injectable drugs. In fact, the number of injections billed per ambulatory visit dropped by 60 percent over the two-year period. Most of this dramatic decrease occurred between January and May 1972 (i.e., after the guidelines were initially distributed and educational efforts begun); the decline continued beyond May (i.e., after enforcement in the form of denials of payment was begun in earnest).

All physician types contributed to this drop in the use of injectable drugs. Outliers improved their injection practices more than any other single physician category.

Although we judged the efforts of the EMCRO to reflect a clinically meaningful improvement in quality of care, the problem was not considered solved. By the end of two years, only about one of every 33 Medicaid eligibles had one injection per month, but fully one-third of those were judged medically inappropriate by the review physicians. Put another way, by the end of the study, about 1 in every 6 ambulatory visits was associated with use of an injectable drug, meaning that approximately one in 20 ambulatory visits still included an inappropriate injection.

SECOND (FOUR-YEAR) STUDY: ALL INJECTABLE DRUGS

The next evaluation of the New Mexico EMCRO focused chiefly on the inpatient hospital setting. One element of that study, however, was a look at the changes in the use of injectable drugs over the entire four years of the EMCRO's operation (Brook et al., 1978, Chapter 2). The main findings were similar to those described above. Some 167,690 injections had been billed and 43,070 (about 26 percent) denied for medical reasons. About half of those were antibiotics: other large classes of injections were tranquilizers, steroids, and analgesics.

Special analyses were done at this time on data from an unchanging cohort of Medicaid eligibles--those who had been enrolled in the Aid to Families with Dependent Children (AFDC) aid category continuously for

the entire four years of the study. Regression analyses indicated that the number of injections billed per 100 ambulatory visits for this cohort decreased 76 percent between the start and end of EMCRO operations. Thus, most of the EMCRO's success in reducing the use of injectable drugs in line with its guidelines occurred in the first two years (indeed, in the first year) and was maintained fairly consistently thereafter.

As with the two-year study, a nontrivial amount of inappropriate injection treatment was observed even at the end of the four-year study. More than 20 percent of all injections were still being denied for medical reasons at the beginning of the fourth year of EMCRO operations, and about 16 percent at the end of that year (Lohr et al., 1980, Chapter 2). The types of injections that apparently continued to be misused by at least some types of physicians included short-acting penicillin, antihistamines, certain kinds of tranquilizers, steroids, and hormones. The greatest declines in the use of injections were observed among those pharmacological groupings with potentially serious side effects, those with more specific indications for use, and those that were used the most when the EMCRO program first began.

The earlier findings about characteristics of physicians associated with better quality of care as measured by use of all injections and about changes over time in use of injections by the various physician categories were observed also in the four-year study (Lohr et al., 1980, Chapter 3). The characteristics related to more appropriate use of injections were group practice (compared with solo practice), especially for the few multi-specialty clinics that contributed the bulk of the

group practice visits, being board certified, being an MD rather than a DO, and being an MD pediatrician (compared with most other specialties). Again, all physician types improved their use of injectable drugs, especially the small set of outlier physicians.

THIRD (TWO-YEAR) STUDY:* INJECTABLE ANTIBIOTICS

To this point, little was known about why injectable drugs had been used so heavily in the Medicaid program. The third phase of the EMCRO studies thus investigated ways in which different types of physicians in private practice chose to treat common infectious illnesses among their Medicaid patients, as a way of clarifying for what diagnoses the largest class of injectable drugs--antibiotics**--were used. The analyses were restricted to antibiotic injections mainly because their use was widespread and represented about half of all injections given during the period of this study, and because the EMCRO diagnosis-specific criteria for antibiotics were more specific and rigorous than criteria for some of the other commonly used injections. In addition, other researchers (Ray et al., 1976, 1977 a,b; Schaffner et al., 1978) have shown that certain antibiotics (in oral form) have been misused in the Tennessee Medicaid program. Further, some antibiotics (including those misused early in the EMCRO's operation) have potentially serious side effects in both adults and children.

*The literature reviews, methods, and results of the third study are described in Lohr et al. (1980); the material on episodes of care is in Chapter VI of that publication. It will not be referred to further in this or succeeding sections.

**In all this work, "antibiotics" is used to refer to both natural and synthetic antimicrobial drugs.

The major findings of this set of analyses included the following: Five injectable drugs (short-acting penicillin, long-acting penicillin, ampicillin, tetracycline, and lincomycin) and about ten diagnoses (acute upper respiratory infection, pharyngitis with tonsillitis, bronchitis, otitis media, diarrhea, cystitis, influenza, skin infections, lacerations, and strep throat) accounted for about 50 percent of the ambulatory visits and nearly 80 percent of the antibiotic injections in Period I. Generally speaking, the same group of diagnoses accounted for roughly the same proportion of ambulatory care visits in the Medicaid population. By Period II, however, the number of antibiotic injections given to treat these types of conditions had decreased markedly. The only diagnosis still managed with injectable antibiotics was strep throat.

In treating common respiratory infections, some types of physicians (classified as before with respect to professional characteristics) exhibited different preferences for specific types of antibiotic injections irrespective of diagnosis. For example, within the diagnoses studied, MDs tended to prefer short-acting penicillin; DOs, especially outliers, relied on lincomycin. These preferences were more marked in Period I than Period II.

One major question was the degree to which diagnoses might have shifted over time, reflecting attempts by physicians to justify use of injectable antibiotics in situations where that use might be questioned and the Medicaid claim denied for payment. The hypotheses were that (1) if physicians were trying to circumvent EMCRO injection guidelines, the ratios of bacterial to viral diagnoses would rise (by Period II) because

bacterial infections treated with injectable antibiotics would be more likely than viral infections to pass the EMCRO guidelines and (2) if physicians were labeling more patients with respiratory infections as a relatively more bacterial diagnosis, the ratio of strep throat to pharyngitis would also rise.

Generally speaking, these two hypotheses were not borne out. With respect to the bacterial/viral distinction, no increase in the percentage of visits classified as bacterial was observed for any physician type except all noncertified DOs (including outliers) and noncertified group practices. The percentage of "sore throat" cases labeled strep throat increased slightly (from 16 to 19 percent between the two periods), but this was viewed, from any clinical or programmatic point of view, as not a particularly meaningful change. This rise was accounted for by noncertified group practices, outlier DOs, and certified DOs. Certified group practices and certified and noncertified MDs were quite consistent over time in their use of the specific diagnosis strep throat. Decreases in the percentage of sore throat cases so labeled were observed for all the remaining provider types.

THIRD (TWO-YEAR) STUDY: EPISODES OF CARE

The objectives of this last phase of the earlier studies were as follows: (1) Determine if entire episodes of care could be used as an unit of analysis in evaluating quality of care; (2) investigate the care provided for episodes of common ambulatory illnesses using a restricted set of process criteria involving antibiotic medications (both injectable and oral) or common laboratory tests; and (3) attempt to

replicate some of the earlier studies, which showed that use of injections differed by physicians characteristics and improved over time.

Using Episodes of Care in Ambulatory Care Review

Recognition is growing that review of ambulatory care should be based on all the care given for an entire illness episode instead of on the care given at a single office visit. This approach theoretically enhances the validity of quality assessment by allowing patterns of care to be considered, permitting investigation of the possible substitution of services, and providing an opportunity to examine the promptness of services and the use of services over time. It is in effect the only feasible way that composite quality-of-care criteria, which simultaneously take into account the presence and/or absence of both good and bad elements of care, can be used.

Although the literature on episodes dates to the late 1960s (Solon et al., 1967; 1969), it is nonetheless sparse. Donabedian (1978) emphasized the need for review of entire episodes, and several researchers have developed a variety of ways for creating episodes of care. None of this major work on episodes of care* simultaneously used insurance claims data and was oriented towards quality-of-care assessment. Consequently, one major effort in the last phase of the earlier studies was to explore how well episodes of care could be

*See Solon et al., 1967, 1969; Lasdon and Sigmann, 1977; Moscovice, 1977; Kane et al., 1976a, b, 1977, 1978; Roos et al., 1977a, b, 1978, 1979; Gold, 1979.

developed from computerized algorithms applied to Medicaid claims data and how well quality of care could be assessed on the basis of those episodes.

The episode studies focused on the AFDC cohort and the previously used physician characteristics--type of practice (group/solo), type of physician training (medical/osteopathic), board certification status, primary care specialty, and outlier status. They were restricted to six common respiratory infections: streptococcal sore throat, pharyngitis and/or tonsillitis, otitis media, acute bronchitis, influenza, and acute upper respiratory infection (URI). The data source was a special tape produced by Dikewood for Periods I and II (defined above). The tape contained the following data items: ambulatory visits; injections, grouped according to 38 separate pharmacologic groups; all prescription drugs, coded by both the national drug code and a special therapeutic code developed Dr. Richard Atkinson of the New Mexico Department of Health and Social Services; laboratory tests and diagnostic procedures that had been selected by the research team for their relevance for these study diagnoses; patient information such as age and sex; diagnosis, coded according to HICDA-I codes; and a provider linking number that specified the solo physician or group practice who gave the particular service.

Simple disease-specific criteria for appropriate use of antibiotic medications in both injectable and oral form were developed; they basically specified what antibiotics would be appropriate for the diagnoses under investigation and situations in which no antibiotic use would be considered acceptable (see Table 6.1 of Lohr et al., 1980). In

addition, a few diagnostic tests were specified as elements of good quality are--for instance, throat or strep cultures for episodes of pharyngitis or strep throat. A physician followup visit was also specified as an element of good quality care for otitis media. The criteria took such factors one at a time in assessing quality. Moreover, the earlier studies did not deal with any other medications such as common symptomatic drugs or with a variety of other laboratory or diagnostic tests that occurred with some frequency in this data set (although not necessarily in episodes of these conditions).

One important step in developing disease-specific episodes based on insurance claims was to specify a length of time that the episode would run. For these respiratory conditions, episodes were considered on clinical grounds likely to last longer than one week but typically not longer than four weeks. Between these extremes, an empirical approach was used to select an appropriate length of time; two weeks was the period finally chosen.

To assign each physician visit, laboratory test, or injection to one and only one of two diagnoses that might appear on a Medicaid claim, a set of ranking rules was established that specified a hierarchy from diagnoses with a clearly bacterial origin to those that were more likely viral in etiology. Whenever a service could be assigned to more than one diagnosis, the more bacterial was used. This was intended to give the benefit of the doubt regarding antibiotic use to the physician.

For the analyses carried out in the last part of the earlier analyses, episodes were constructed according to a set of rules summarized as follows. For each patient an episode was considered to

begin with a visit for one of the study diagnoses and to continue for two weeks thereafter. All visits, laboratory tests, injections, and prescription drugs occurring in the relevant time frame were assigned either to the diagnosis under study (i.e., that of the initial visit) or to one of two other diagnostic categories: "closely related" diagnoses or "completely unrelated" diagnoses. Closely related diagnoses were defined as any of the other five study diagnoses or a set of other relatively common infectious diseases of the respiratory tract (sinusitis, bacterial pneumonia, laryngitis, and tracheitis); completely unrelated diagnoses were all remaining diagnoses. To avoid partial episodes, episodes were constrained to begin at least two weeks after the start of Periods I or II (or two weeks after the date of a visit for one of the study diagnoses if that visit occurred within the first two weeks of either period), and not to begin within two weeks of the end of each period. A set of rules was developed to handle combinations of both oral and injectable antibiotics (separately and together); these rules resulted in assigning each episode to one and only one of 12 antibiotic injection categories and one and only one of 16 oral antibiotic categories, and then to one of 192 (12 times 16) possible combinations, according to the injectable or oral antibiotics observed in each episode.

One major complication to the episode work is that diagnoses are not specified on prescription (pharmacy) claim forms in the Medicaid program. To be able to assign prescription drugs to a specific diagnosis, therefore, an algorithm needed to be developed by which this could be reliably done. To this end, a set of rules generically

referred to as the "prescription rule" was developed such that, within the time frame of the episode, a prescription drug was assigned to the diagnosis of the first (initiating) visit in the episode if no other visit intervened between the date of the first visit and that of the prescription. The prescription was assigned to the diagnosis of an intervening visit that if that diagnosis was different from the initial one.

Certain implications of the episode rules should perhaps be noted here. First, services or medications with a date of delivery before a visit beginning an episode were simply not included in the episode; similarly, services that might belong to the episode but that occurred after the fourteenth day were also not included in it. Third, a variety of methodologic questions might be raised about using the hierarchical ranking to choose between two diagnoses on a given claim, and about dealing with complex episodes containing several visits or other services with competing respiratory infection diagnoses. These kinds of questions were not systematically addressed in the earlier studies, and are therefore the focus of specific investigation in this study (particularly as reported in Chapter II on the validation studies).

Results of the Initial Episode Analyses

Only the highlights of the results of the original analyses are recapitulated here. They focus chiefly on: (1) The differences in quality of care by provider types (where provider type is as previously defined) and the differences in quality-of-care between Periods I and II; (2) special analyses involving use of tetracycline in children under

8 years of age; (3) use of multiple pharmacological classes of antibiotics ("polypharmacy") and of lincomycin; (4) differences among provider types and between the two periods of use of selected diagnostic procedures; and (5) use of followup visits for otitis media.

The first hypothesis tested was that quality scores would be higher in Period II than I; findings were all consistent with that hypothesis. Generally, the percentages of episodes treated appropriately (taking all provider types together) ranged fairly widely; in Period II, the overall range had shifted upward compared with Period I, such that between 42 percent (strep throat) and 81 percent (bronchitis) of episodes were treated correctly. In this instance, treated correctly refers strictly to the use of appropriate antibiotics (or no antibiotics) for the six study conditions. Overall, however, only about one half of all episodes of these common diseases were being handled appropriately even in Period II, and only bronchitis showed anything like what might have been hoped for in the form of achievable improvement in use of antibiotic drugs. Moreover, these findings were derived from an evaluation based on criteria for a minimal level of quality of care, which suggested that the level of basic care encountered in the Medicaid population was still in need of improvement.

The second hypothesis was that group practices, taken collectively, would give better care (again as measured by use of proper antibiotics or no antibiotics) than physicians in solo practice (as had been shown in the various earlier studies). This second hypothesis was confirmed in virtually every disease and age grouping in both periods.

A third hypothesis was that the "outlier" physicians would give the poorest levels of care as judged by these antibiotic criteria; this, too, was borne out. Because the outliers generally had been observed to decrease their use of injectable antibiotics substantially, we concluded that their persistent deficiency in quality was attributable to a poor choice of oral medications, some unacceptable "polypharmacy," and some continuing overtreatment, largely of viral conditions.

With respect to other physician characteristics, findings were mixed. Generally but not always, MDs had a higher proportion of adequately treated episodes of all conditions than did DOs; MD general practitioners gave significantly better care than DO general practitioners. Typically, board-certified physicians within both the MD and DO categories gave better care than noncertified physicians across all diseases. Overall trends supported an hypothesis of better quality of care among board-certified physicians, especially for MDs.

The greatest improvement in use of antibiotics for these diseases was observed for physicians who generally provided the poorer levels of care in Period I. That is, results tended to support the hypothesis that the most positive changes between Periods I and II would be observed among the poorer performers initially, despite any finding that care given by such physicians might remain at unacceptably low levels in Period II. Because of these greater improvements among the poorer performers without any concomitant decrements among better providers, we argued that the variance among providers was substantially reduced between Periods I and II.

For the more descriptive analyses involving diagnostic tests and physician visits, the following findings were reported: Use of highly

relevant (or recommended) laboratory tests such as throat or strep cultures for strep throat or pharyngitis was extremely low in both Periods I and II. Some small increase in the use of at least one relevant procedure (e.g., throat culture or white blood cell count) was observed over time, but no inference could be drawn about any causal effect of EMCRO operations, because the EMCRO did not systematically set out to alter physician behaviors in this regard. Generally, patterns of use of laboratory tests by physician type were analogous to those for antibiotic use, namely, higher use of relevant tests by group practices and virtually no use by outliers. When quality of care was evaluated in terms of revisit rates for otitis media episodes, less than one-third of all episodes has a required followup visit, and use of followup visits in otitis media differed less by provider type than did the use of antibiotics or lab tests.

Comment on the Initial Episode Analyses

Drawbacks to the initial episode analyses can be summarized as the following:

(1) Comparatively little validation of the initial episode rules was done at that time, leading to some ambiguity about the rigorousness of this new methodology. These types of questions are taken up in the validation studies (Chapter II).

(2) Although an effort was made to develop quality-of-care criteria involving the joint use of injectable and/or oral antibiotics, no comprehensive set of criteria that takes into account other medications

relevant (or recommended) laboratory tests such as throat or strep cultures for strep throat or pharyngitis was extremely low in both Periods I and II. Some small increase in the use of at least one relevant procedure (e.g., throat culture or white blood cell count) was observed over time, but no inference could be drawn about any causal effect of EMCRO operations, because the EMCRO did not systematically set out to alter physician behaviors in this regard. Generally, patterns of use of laboratory tests by physician type were analogous to those for antibiotic use, namely, higher use of relevant tests by group practices and virtually no use by outliers. When quality of care was evaluated in terms of revisit rates for otitis media episodes, less than one-third of all episodes has a required followup visit, and use of followup visits in otitis media differed less by provider type than did the use of antibiotics or lab tests.

Comment on the Initial Episode Analyses

Drawbacks to the initial episode analyses can be summarized as the following:

- (1) Comparatively little validation of the initial episode rules was done at that time, leading to some ambiguity about the rigorousness of this new methodology. These types of questions are taken up in the validation studies (Chapter II).
- (2) Although an effort was made to develop quality-of-care criteria involving the joint use of injectable and/or oral antibiotics, no comprehensive set of criteria that takes into account other medications

and/or diagnostic tests or various other procedures was used. Some findings with regard to use of antibiotics (especially in strep throat) were interpreted in the light of findings regarding the use of laboratory tests; in particular, somewhat different criteria were used for physician types whose use of throat cultures in strep throat exceeded a certain threshold and where the assumption could be made that a throat culture had been used to rule out strep throat and a negative result had led to no antibiotics being prescribed or injected. By and large, however, no effort had been made to use any type of composite criteria. Consequently, the Profiles of quality of care (Chapter III) were designed in part to meet this need for a more comprehensive and unified set of criteria.

(3) No effort had been made in the previous studies to evaluate quality of care, and improvements over time in that care, in light of the cost of care. Estimated dollar "savings" for all injections prevented by EMCRO activities (taking injections of all types of medications per ambulatory visit in the Medicaid population as the metric), in comparison with the estimated program costs of the EMCRO's quality-of-care (peer review) activities, were at best negligible. We argued at the time, however, that there is no particular programmatic or social reason for insisting that quality-of-care review should "pay its own way" in situations of this sort, when the presumed benefits to patients individually and to society generally cannot easily be measured in dollar terms and certainly cannot be measured exclusively as the price of services foregone. Assessing this "quality/cost tradeoff," which has received wide publicity over the years, is also not done in

the present study. Although such analyses are sufficiently important to warrant investigation in the future, they were beyond the scope of the present report.

CHAPTER II
VALIDATION OF EPISODE RULES

The validation studies were intended to determine whether the original "New Mexico" rules for assigning services to respiratory disease episodes in the New Mexico Medicaid claims data base were clinically valid and logically consistent, whether any systematic errors were introduced by those rules, and whether they could be improved. The ultimate objective was to arrive at the best possible algorithms for defining episodes of care in the New Mexico data base, so as to enhance the validity of the subsequent quality-of-care analyses. Consequently, the tasks undertaken included not only testing the original rules and proposing modifications as necessary but also reviewing difficult-to-classify episodes in an attempt to uncover any additional revisions that might be warranted.

No medical record or other data were available from New Mexico against which to validate the New Mexico episode algorithms directly. For this reason, external data were sought that had two main characteristics. First, they should allow "true" episodes to be defined without extensive recourse to arbitrary or a priori rules. Second, they should be as accurate, complete, and free from methodologic bias as possible. A third factor was that using insurance claims data was considered preferable to using other types of data (e.g., medical records) to minimize any problems that might arise from using two different types of data for the validation study.

The basic approach was to apply the New Mexico episode rules to the external data and compare the episodes so generated with the "true" episodes in those data. The degree and type of differences between the two sets of episodes would indicate problems with the rules (if any) that would need to be corrected to yield more valid episodes. The revised rules would then be applied to the external data and any observed improvements (in the sense of fewer differences) would be noted.

The remaining sections of this chapter are organized as follows: (1) a description of the external insurance claim data from the Rand Health Insurance Study (HIS); (2) methods of defining HIS episodes; (3) results of special analyses of HIS episodes; (4) methods of applying the New Mexico algorithms to the HIS episodes; (5) the results of the validation analyses; and (6) a discussion of the nature, importance, and implications of certain "errors" and complications in the episode methodology as applied to claims data.

RAND HEALTH INSURANCE STUDY DATA

The external data were taken from ambulatory care insurance claims from the Rand Health Insurance Study (HIS). In particular, all available information from complete "patient profiles" for the first HIS site (Dayton, Ohio) for the first two years of the HIS enrollment period

was used.* The HIS data were considered eminently suitable for this validation work for several reasons, which are noted below and discussed in more detail in Appendix A.

First, data on ambulatory services are compiled from specially designed HIS claim forms (Figure II.1) that are considerably more comprehensive than the usual insurance claim. In particular, the HIS "Physician Outpatient" claim has a problem-oriented format that requires the physician to record one or more diagnoses/problems and to relate services/procedures to these diagnoses/problems. Second, up to four spaces on the "Physician Outpatient" claim are allowed for diagnoses/problems to be listed by the physician. This gives both the physician and the researcher greater flexibility in linking and analysing services and medications by diagnosis.

Third, prescriptions for medications can be linked more easily into episodes of care with the HIS data than with other types of claims data, even if some information, including diagnosis, is missing. Physicians who actually fill their own prescriptions or provide medications directly to the patient can bill for those prescriptions on the Physician Outpatient claim form. Those who write prescriptions to be filled elsewhere are still required to record the prescription, complete with diagnosis, on the Physician Outpatient claim. Typically, a good deal of confidence can be placed in any match between a prescription recorded on the Physician Outpatient claim and the prescription billed on the Pharmacy claim.

*A more complete description of the Rand Health Insurance Study can be found in Newhouse (1974) and Ware et al. (1979).

FIGURE II.1

W84450

INSTRUCTIONS: 18002
Participant must complete Items 1 - 14. Before giving the form to the Doctor, tear off Visit Record, complete questions on the back, and keep for your records to fill out your Health Report. Use this form for all doctor visits, clinics, emergency, outpatient hospital surgery, supplies, and x-ray and laboratory charges.

FAMILY HEALTH PROTECTION PLAN

PHYSICIANS, DOCTORS, SUPPLIERS AND OUTPATIENT MEDICAL EXPENSE REPORT

(Use this form for all outpatient charges clinics, surgery, emergency, etc.)

PART 1 PARTICIPANT TO FILL IN ITEMS 1 THROUGH 14 PLEASE PRINT OR TYPE					
1. Last Name of Patient	First	M.J.	2. Sex	3. Age	4. Patient's Family No.
5. Patient's Address	City, State, Zip Code			6. Patient's Individual No.	
7. What Was The Major Reason or Symptom For This Visit To The Doctor? (For illness or Accident)		8. Was Illness or Injury Employment Related?		9. Was Illness or Injury Accident Related?	
<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO		<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO		<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO	
				10. Date of Injury or Accident 	
11. Describe how and where accident occurred					
12. Name of Doctor, Supplier or Outpatient Facility					
13. Has the Patient Ever Visited This Doctor, Supplier or Outpatient Facility Before? <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO					
14. I authorize any holder of medical or other information about the patient to release to the Family Health Protection Plan or its intermediaries any information needed for the or related medical reports I permit a copy of this authorization to be used in place of the original in conformance with the Family Health Protection Plan Establishment Agreement, all health care benefits covering the Patient are hereby assigned to the Family Health Protection Plan.					
Signature of adult Participant or Guardian of Minor Participant			Print Adult's Name		Date Signed

PART 2 DOCTOR OR SUPPLIER TO FILL IN ITEMS 15 THROUGH 29 PLEASE PRINT OR TYPE					
15. Full Name of Referring Doctor (If None, Write None)	16. Full Name(s) of Providers to Whom You Referred Patient for Consultation, Lab Tests, or Other Services (If None, Write None)				
17. Describe the Primary Problem or Diagnosis Which Brought the Patient To Your Office and Any Other Problem(s) for Which You Supplied Treatment. Please List Primary Problem or Diagnosis on Line A			18. Type of Problem (check one)		19. Treatment History (check if well care or pregnancy)
A. B. C. D.			<input type="checkbox"/> Acute <input type="checkbox"/> Well Care <input type="checkbox"/> (or pregnancy)		<input type="checkbox"/> Flare-up of Chronic <input type="checkbox"/> Chronic (not Flare-up)
			<input type="checkbox"/> Acute <input type="checkbox"/> Well Care <input type="checkbox"/> (or pregnancy)		<input type="checkbox"/> Flare-up of Chronic <input type="checkbox"/> Chronic (not Flare-up)
			<input type="checkbox"/> Acute <input type="checkbox"/> Well Care <input type="checkbox"/> (or pregnancy)		<input type="checkbox"/> Flare-up of Chronic <input type="checkbox"/> Chronic (not Flare-up)
			<input type="checkbox"/> Acute <input type="checkbox"/> Well Care <input type="checkbox"/> (or pregnancy)		<input type="checkbox"/> Flare-up of Chronic <input type="checkbox"/> Chronic (not Flare-up)

KEY
 Place of Service Codes: O = Doctor's Office IL = Independent Laboratory H = Patient's Home IH = Inpatient Hospital NH = Nursing Home or SNF ER = Emergency Area
 OH = Outpatient Hospital including Hospital Clinic and Outpatient Surgery SC = School Clinic CC = Company Clinic OL = Other Location including Other Non-Hospital Clinic
 Type of Visit Codes: 1 = Minimal Service 2 = Brief Examination 3 = Limited Examination 4 = Intermediate Examination 5 = Extended Examination
 6 = Comprehensive Examination 7 = Unusually Complex Examination
 For Inpatient Services, Omit 18, 19 and 21
 SEE DETAILED INSTRUCTIONS ON REVERSE SIDE

A	B	C	D	E	F	G
Date of Service	Place of Service Use code above	Describe Each Medical or Surgical Procedure and Other Service or Supplies Furnished For Each Date including Test(s), Lab Test and the Specific Name of Any Drug Required.	Type of Office Visit Use code above	Reason Treatment is Given by Ref to 18 & C or D above	Charge	
1						
2						
3						
4						
5						

22. Name and Address of Doctor or Supplier	23. Social Security or Provider Tax ID Number	24. TOTAL CHARGE	
	Telephone Number	25. AMOUNT PAID IF ANY	
		26. BALANCE DUE	

27. I hereby certify that the services and/or supplies noted above have been provided on the date(s) and in the amount(s) indicated.
PROVIDER'S SIGNATURE
 28. I hereby authorize payment directly to the above-named provider of the benefits otherwise payable to me but not to exceed the charges shown. I understand that I am financially responsible for any charges not covered by the Family Health Protection Plan.

ADULT PARTICIPANT'S SIGNATURE
 Date Signed

NME 6871 REV 1-77 12M 11-79

MAIL TO FHPP

THIS PAGE IS BEST QUALITY COPY FURNISHED TO DDC

METHODS OF ABSTRACTING HIS EPISODES

The first steps were to extract manually all relevant episodes from the HIS computer-generated patient profiles and to record all information pertinent to these analyses on a special form. The forms and methods are described in more detail in Appendix A.

All episodes with physician visits bearing the diagnoses pharyngitis (HICDA-II 462), tonsillitis (463), and streptococcal sore throat (34.0), or the symptom "sore throat" (777.6) were used. They are referred to generically as sore throat episodes in this chapter.

Pharyngitis and tonsillitis are combined into a single disease entity (referred to as pharyngitis); the symptom "sore throat" is referred to as a diagnosis simply for ease of presentation.

This spectrum of sore throat diagnoses was chosen because it could be expected to yield a sufficient number of cases and because it represented a range of seriousness of the condition (from strep throat [a bacterial condition probably needing antibiotic therapy] to "sore throat" [a somewhat nebulous symptomatic designation not suggesting any pathologic agent and probably not needing antibiotic therapy]). No distinction was made among the four diagnoses, because doing so would have served little analytic purpose and would have yielded extremely small sample sizes for strep throat and "sore throat."

An episode was identified initially by a Physician Outpatient claim with a visit bearing one of these four diagnoses. The "triggering" diagnosis could be the first, second, third, or fourth diagnosis that had been recorded by the physician on the claim.

Information from the profiles that was transferred to a special form included the following: "ID" numbers for the episode and the patient; dates of service and diagnostic codes; all services as of the date of a visit beginning an episode; related services before the visit date and up to 3 or 4 months after it (where related refers to services with one of the validation-study diagnoses); all other services during or following the episode; and data on prescriptions written, dispensed, and filled. Details can be found in Appendix A.

SPLITTING THE HIS EPISODE SAMPLE

The basic purpose for these studies, as noted above, was to ascertain whether the original New Mexico rules adequately defined episodes of respiratory infections, by comparing "true" episodes from an external source with episodes from that same source defined by the New Mexico rules. In the event that discrepancies were observed, an attempt would be made to ascertain how the New Mexico rules were deficient (in the sense of missing or misassigning services to episodes). Based on that analysis, modifications to the rules would be made to correct the problems. Because modifying the original rules might lead to other problems, it was decided to do the first analyses on only half of the "true" episodes, so that modified rules, if any were needed, could be tested on the other half of the "true" episodes. (This is the conceptual equivalent of modeling on one-half of a data set and testing hypotheses about that model on the other half.)

To this end, once HIS episodes were abstracted from the patient profiles to the special form, they were divided such that "Group A" (as

it will be called hereafter) comprised every odd-numbered episode (i.e., every other episode beginning with number 1) and "Group B" comprised every even-numbered episode. These two groups did not differ with respect to level of complexity, defined as the number and percent of episodes that had (1) only visits; (2) visits and medications; (3) visits and laboratory work; or (4) visits, medications, and laboratory work.

ISSUES IN ABSTRACTING HIS SORE THROAT EPISODES

Multiple Diagnoses

Defining "true" HIS episodes required, first, that diagnoses either be present or be imputed for all visits, services, and medications that might possibly belong to the episode. HIS episodes can be divided into those in which the physician had specified unambiguous diagnoses for all visits, procedures, or medications (injections, prescriptions written, or drugs sold or dispensed), and those in which some problem arose because of ambiguity as to the diagnoses intended or missing diagnoses. These points are taken up in turn below. (In the discussion below, "services" is used broadly to refer to visits, tests, procedures, and injectable and oral drugs.)

The simplest and most frequently used approach to define an HIS episode was to use the fact that, for some patients, all services within some relatively discrete period of time (e.g., up to six months or so) had only one unchanging sore throat diagnosis recorded on the claim

forms.* One variation on this direct approach was that for some patients, most services had the same single sore throat diagnosis; the remaining services had a diagnosis that was regarded as completely unrelated to sore throat (general medical examination, malaise, hypertension, etc.). A second variation was that some or all services had a "pair" of diagnoses consisting of one sore throat diagnosis and one totally unrelated diagnoses; generally this occurred on an initial visit. Altogether, over 70 percent of the HIS episodes had respiratory infection diagnoses that were, for the purpose of this work, unambiguous.

In the remaining episodes, the physician listed either different diagnoses for different services, or more than one diagnosis for some or all of the services that might belong to the episode. This situation involved either (1) two sore throat diagnoses together (e.g., pharyngitis and otitis media), or (2) one sore throat diagnosis and one other respiratory infection diagnosis or symptom (e.g., influenza and cough). For the initial abstracting of HIS episodes, this was handled as follows: (1) Include in the episode all services with only the sore throat diagnosis; (2) exclude services with only any other diagnoses; and (3) include services or medications with "pairs" of diagnoses in which a sore throat diagnosis was one of the elements of the pair.

*This is not to say that the services necessarily extended over the entire time frame, but rather that, for whatever services were observed within the time frame, only one diagnosis appeared.

Missing Data

A second area of difficulty arose when diagnoses were lacking because (1) data on a Physician Outpatient claim simply were missing or (2) oral medications were billed on Pharmacy claims and not recorded on Physician Outpatient claims, in which case no diagnosis was available (from the Physician Outpatient claim) to associate with the medication on the Pharmacy claim.

The "missing diagnosis" problem was handled as follows: If services such as throat cultures and injectable medications (both of which were billed on a Physician Outpatient claim) had no diagnosis coded in the patient profile, it was assumed that they were given for the diagnosis (or diagnoses) associated with the physician visit appearing on the same Physician Outpatient claim.

Decisions about prescriptions (oral) medications with no diagnosis were made in different ways, depending on other data available. If the Pharmacy claim had the same date of service as the date of a Physician Outpatient claim for a visit for one of the sore throat diagnosis, or if the medication had been given previously to the same patient for one of the sore throat diagnoses, then the medication was given the relevant sore throat diagnosis.

The only exceptions to these decisions were for chronic disease medications, which were presumed not to have been given for the sore throat diagnosis. For complex episodes in which such problems occurred, a search was made for many months before the episode to determine the initial reason for the medication.

Table II.1

DIAGNOSIS CHARACTERISTICS OF "SORE THROAT" EPISODES
 FROM THE RAND HEALTH INSURANCE STUDY:
 SINGLE DIAGNOSES

Characteristics of Episode with Single Diagnosis	Number of Episodes	Percent of All Sore Throat Episodes (N=421) ^a
<u>Single Diagnosis</u>		
Strep Throat	51	12
Pharyngitis/Tonsillitis	185	44
"Sore Throat"	<u>44</u>	<u>10</u>
Total	280	66
<u>Single Diagnosis plus totally unrelated diagnosis</u>		
Strep Throat	11	3
Pharyngitis/Tonsillitis	24	6
"Sore Throat"	<u>4</u>	<u>1</u>
Total	39	10
<u>Total Single Diagnosis^b</u>		
Strep Throat	62	15
Pharyngitis/Tonsillitis	209	50
"Sore Throat"	<u>48</u>	<u>11</u>
Total	319	76

^a The total number of episodes is 421, once the various guidelines and modifications to the New Mexico rules are used.

^b Total Single Diagnosis is the sum of the previous two sets of entries in the table, and means that the episode has either just one sore throat diagnosis for all visits, services, and medications, or only one sore throat diagnosis plus a completely unrelated one. Thus, for example, of 421 sore throat episodes, 62 (or 15 percent) were either just strep throat or strep throat plus unrelated diagnoses.

RESULTS OF ABSTRACTING HIS SORE THROAT EPISODESMultiple Diagnoses

Of the final 421 HIS sore throat episodes, 319 (76 percent) had essentially one unambiguous diagnosis by which to assign services (see Table II.1). Some 280 episodes (66 percent) had precisely one diagnosis; this means that all visits, medications, and tests included in the episode had had one and only one (and the same) diagnosis. Another 39 (10 percent) had one of the four sore throat diagnoses plus another completely unrelated one (e.g., hypertension, lacerations and contusions, or general medical examination).

The remaining 102 episodes (24 percent) had "double" diagnoses (Table II.2). This phenomenon was defined as (a) two of the four sore throat diagnoses (typically at the initial visit) or (b) one of the four plus a closely related respiratory diagnosis or symptom. "Closely related" comprised otitis media, acute bronchitis, influenza, acute URI, bacterial pneumonia, sinusitis, laryngitis and tracheitis, hypertrophy of tonsils, and cough.

Of the 421 sore throat episodes, 80 (19 percent) (78 percent of the 102 "double diagnosis" episodes) involved "combinations"--i.e., double diagnoses specified for the index (initial) visit (Table II.2). Of these 80, 66 were strep throat or pharyngitis with each other or with one of the other respiratory diagnoses just named, and 14 were "sore throat" either with cough or with respiratory diagnoses other than strep throat or pharyngitis/tonsillitis (which would already be accounted for). The largest single combination (15 episodes, or 4 percent of 421) involved pharyngitis with otitis media; the next largest (9 episodes, or 2 percent) was strep throat with pharyngitis.

Table II.2

DIAGNOSIS CHARACTERISTICS OF "SORE THROAT" EPISODES
 FROM THE RAND HEALTH INSURANCE STUDY:
 DOUBLE DIAGNOSES

Characteristic of Episode with Double Diagnoses	Number of Episodes	Percent of All Sore Throat Episodes (N=421) ^a
<u>Combinations at index visit</u>		
Strep Throat and all other respiratory infections ^b	7	2
Strep Throat and P/T ^c	9	2
P/T and Otitis Media	15	4
P/T and all other respiratory infections ^d	<u>35</u>	<u>8</u>
Total	66	16
"Sore Throat" and any other respiratory symptoms or infection ^e	<u>14</u>	<u>3</u>
Total Combinations	80	19
<u>Progressions/Changes</u>		
Strep Throat to any other respiratory infection	3	1
P/T to Strep Throat	5	1
P/T to any other respiratory infection ^f	9	2
Another respiratory infection to Strep Throat or P/T	<u>5</u>	<u>1</u>
Total Progressions	22	<u>5</u>
Total "Double Diagnosis"	102	24

^aThe total number of episodes is 421, once the various guidelines and modifications to the New Mexico rules are used.

^bExcept Pharyngitis/Tonsillitis.

^cP/T is Pharyngitis/Tonsillitis.

^dExcept Strep Throat and Otitis Media.

^eExcept Strep Throat and P/T, which, if they occur, are counted in a previous category.

^fExcept Strep Throat.

An additional 22 episodes (5 percent) (22 percent of the 102 "double diagnosis" episodes) involved "progressions"--namely, a change of diagnosis in the episode from one respiratory condition to another (see Table II.2).* Fourteen involved pharyngitis changing to something else (5 cases of strep throat, 3 of bronchitis, 2 each of hypertrophy of tonsils and "sore throat," and 1 each of otitis media and acute URI) and 3 involved an original diagnosis of strep throat changing to something else (1 case of acute URI, 2 cases of otitis media).

Because no dominant patterns of episodes with double diagnoses emerged, the potential complications inherent in such episodes were disregarded for the validation studies. For quality-of-care assessment, however, greater account may need to be taken of both combinations and progressions of diagnoses within episodes. When the diagnoses are not equally "bacterial" (or do not reflect the same bacterium as a likely causal agent) and when the progression is from a less to a more bacterial (i.e., more serious) entity, the chances are that two disease processes are at work simultaneously or that a complication (which may or may not have been preventable) has set in. In such instances, quality of care should perhaps be judged with respect to the more serious condition, and these points are taken up in the discussion below.

*Of the episodes with the diagnosis "sore throat," only one involved a progression, from "sore throat" to "cough." For simplicity, this episode was included with the combinations discussed above.

Missing Data

One other analysis was done to determine whether (or how much) missing data, in the form of diagnoses' not being available, might have distorted the "truth" of the HIS episodes. Diagnoses for laboratory tests, procedures, or injectable drugs were lacking only in very few instances; in most cases when this occurred, the service was billed on the same day (and same claim) as the visit and thus the visit's diagnosis(es) could be ascribed to the test.

Of the final 421 HIS episodes, the diagnosis was present for all medications in the episode or for all medications of interest (i.e., those to be included in the quality-of-care analyses) in 274 episodes (65 percent). In an additional 88 episodes (21 percent), either no prescription was given or a prescription was given but was not filled. Thus, drugs had diagnoses present in 86 percent of the HIS episodes used for the validation work.*

This leaves 59 episodes (14 percent) in which a diagnosis was unavailable for at least one prescription that would be important to any quality assessment (diagnoses may have been available for some but not

*Diagnoses were considered present if one could be assigned using one of the two following techniques: (1) The physician wrote a prescription, recorded it on the Physician Outpatient claim, provided a diagnosis, and the prescription was later filled (as evidenced by one or more Pharmacy claims for that drug); or (2) the physician actually provided the drug and billed for it directly on the Physician Outpatient claim, which also required a diagnosis. In some of these 274 cases, diagnoses were available for the medications relevant to the diagnosis-specific quality-of-care studies but not for other medications irrelevant to those studies (e.g., tranquilizers, antihypertensives, anthelmintics, dermatologic agents, insulin, or vitamins).

all prescriptions in the episode). These 59 involved a total of 86 medications, about which a more or less "arbitrary" decision had to be made. Of these 86 prescriptions, nearly half (37, or 43 percent) were antibiotics, chiefly penicillin, erythromycin, or tetracycline. The next largest group (23 medications, or 27 percent) were cold remedies of various types, including antihistamines (Actifed, for example) and decongestants (Dimetapp, for example). Antitussives and expectorants (chiefly Phenergan and Tussinex) made up 17 of the medications with no diagnosis. The remainder were 5 mild analgesics, 2 bronchial dilators, and 2 otic (ear) drops.

In nearly all the above cases, the decision was made that the prescription belonged to the episode diagnosis, because of proximity in time, plausibility of these therapies' being used for these sore throat diagnoses, and absence of competing or more plausible diagnoses. The exceptions (in the sense that the medications were not attributed to the episode diagnosis) included the following: sulfonamide in an episode involving kidney infection, otic drops in episodes involving otitis media, and bronchial dilators in episodes involving bronchitis. The very few questionable medications included antihistamines in an episode also involving hay fever and tetracycline in an episode that was followed shortly by a hospitalization and abortion.

The above discussion focused on episodes, and one missing data item (i.e., diagnosis) was counted as a "miss," regardless of the number of items in the episode. This potentially overestimates the true "missing diagnosis" rate. To explore this further, a 10 percent random sample of episodes was investigated. The number of services in these episodes was

counted, and the number with a missing diagnosis noted. A distinction was made between services relevant to the sore throat conditions and those not relevant, as they had been recorded from the patient profiles onto the special form.

The 41 randomly selected episodes had a total of 128 relevant services and 29 irrelevant services. Of the 128 services (all drugs), 12 (9 percent) were missing a diagnosis. These included 5 instances of antibiotics; the remainder were antitussives, decongestants, or expectorants. With one exception (an antibiotic that might have been given for a kidney disease diagnosis), these medications are all highly likely to have been given for the sore throat diagnosis. Of the 29 "irrelevant" services, 18 (64 percent) had no diagnosis recorded from the patient profile. Of these 18, 9 were in one episode and involved repeated injections of compazine and a placebo and a prescription for paragoric; another 4 were in one episode and involved a sulfonamide typically used for urinary tract infection; the remaining 5 were an antihypertensive, an anti-asthmatic, a tranquilizer, a steroid cream, and an antidiarrheal. The high rate of missing diagnoses for these "irrelevant" medications is somewhat misleading; a re-review of the original patient profile indicated that many were given over a very long period and that a diagnosis would have been available (often months earlier).

In general, then, the level of doubt about the diagnosis associated with prescription drugs used in these HIS sore throat episodes was very low. Using the number of services in the episodes as the metric, less than 10 percent were missing a diagnosis, and less than 1 percent were

questionable in the sense that a competing diagnosis was a possible choice. Although some "arbitrary" decisions had to be made about the diagnosis to impute to a prescription lacking a diagnosis, in most cases a good deal of confidence could be placed in the decision to impute a sore throat diagnosis to the drug and include it in the episode (because the type of drug was so likely to be used in precisely these conditions) or to impute some other diagnosis to it and exclude it from the episode (because a more "logical" diagnosis had been recorded for another service [chiefly visits]). The conclusion, then, was that the HIS episodes were adequately "true" for the purposes of the validation studies.

Very Long or Complex Episodes

Initially a total of 408 episode were abstracted from the HIS patient profiles to the special form (i.e., 204 each in Groups A and B). Some of the Group A episodes were fairly complex and difficult to define unambiguously from the patient profile. Some of the Group B episodes, similarly, were lengthy and/or complicated. In keeping with the notion of "modeling" on the Group A episodes and "testing" on Group B, the following analysis was restricted to Group A.

A board-certified internist was asked to review 35 complicated Group A episodes and to make a medical judgment as to the most likely clinical configuration of the episode. In instances where services occurred several weeks after the initial visit for the acute condition, the clinician was asked to indicate whether they belonged to that episode or constituted a second (or third) episode. The following clinically oriented guidelines emerged.

(1) Lengthy episodes (of, say, six weeks) in which the bulk of the services were not for one of the four sore throat diagnoses (e.g., episodes in which most services were for bronchitis or kidney disease) were dropped from the validation analyses because they were more likely to be a condition other than the ones of interest. (2) Episodes involving infectious mononucleosis were dropped from the validation analyses, on the grounds that in all likelihood the clinical problem was not of interest in this study. (3) To distinguish between "long" episodes and second or subsequent episodes of the same diagnosis, the following guidelines were proposed: (a) If a visit for the same diagnosis occurred within 18 days of the index visit (i.e., no more than three days after the end of a "two-week" episode), and if that visit was associated with an injectable or oral antibiotic, it began a new episode; (b) if a visit (or other services) for the same diagnosis occurred within 18 days of the index visit but was not associated with any antibiotic, it belonged to the original episode.* (4) A visit for the same diagnosis occurring more than three days after the end of a two-week episode began a new episode. (5) A visit for a different diagnosis occurring after two weeks either began a new episode (for that different diagnosis) or was ignored (if it was not a sore throat diagnosis). These "clinical" guidelines were eventually used in both Groups A and B, as described more fully below.

*These two guidelines were invoked in very few of these 35 episodes. In most cases, visits occurring after the end of the episode occurred well more than three days later (in fact, more than a week or two) and were thus considered to begin a new episode.

VALIDATION STUDIESMETHODS OF APPLYING THE NEW MEXICO RULES TO HIS EPISODES

Several weeks after the HIS episodes had been abstracted from the patient profiles and the other analyses done as described above, the New Mexico rules for creating episodes were used to "re-create" a set of episodes. (The reader is referred to Chapter I for a description of the original New Mexico rules used in the previous study.) This step was done initially on just the Group A HIS episodes. The outcome of applying the New Mexico rules to each HIS episode was recorded, and all discrepancies between the elements in or characteristics of the "true" HIS episode and the elements or characteristics of the "New Mexico" episodes were explicitly noted. For example, if the New Mexico rules would have yielded an episode with exactly the same characteristics as the "true" HIS episode had, it was considered an identical match.

Differences between the HIS episode and those created by applying the New Mexico rules could be of several types: (1) Missing services because they appeared before the index visit that begins an episode and determines its diagnosis; (2) missing services for the diagnosis of the episode that were given more than two weeks after the index visit; (3) wrongly including services not truly given for the diagnosis; (4) wrongly excluding services that were given for the diagnosis; and (5) various combinations of these four errors. In comparing the true HIS episodes with those created by the New Mexico rules, these types of errors (or others) would be explicitly noted.

The types of errors in the Group A episodes were examined; based on those findings, various revisions to the episode algorithms were proposed. As implied by the above description of methods for defining HIS episodes, these revisions took two forms. One set of proposed modifications to the rules was intended to simplify very long and complex episodes in which the true course of illness would not be clear even from the data in the patient profiles; the other set of modifications was designed to reduce errors that seemingly arose from the New Mexico algorithms themselves.

The "original" New Mexico algorithms and the various "new" algorithms were then applied to the Group B HIS episodes, to see what improvements (if any) they made, in the sense of fewer services being incorrectly included or excluded from the episode. These steps and the outcome of these analyses are described below.

RESULTS OF APPLYING THE NEW MEXICO RULES TO HIS EPISODES

Group A Episodes

Initially, a total of 408 HIS episodes were abstracted, 204 each in Groups A and B. Of those in Group A, 160 (78 percent) created by the New Mexico rules were identical to the "true" HIS episodes. That is, applying the New Mexico computer algorithms to the HIS data gave episodes with all elements identical to those in the HIS episodes.

The remaining 44 episodes (22 percent) were not identical to the HIS episodes. Of these, 7 (3 percent) differed from the HIS episodes only in that the New Mexico rules would have included services for a closely related respiratory illness (e.g., services for URI in an

episode classified as sore throat). An additional 37 episodes (18 percent) had other differences from the HIS episodes, for a total of 42 "errors." These are discussed in more detail below.

Episode Length

In these original Group A episodes, three main differences were observed between the HIS episodes and those defined by the New Mexico rules (see Table II.3): (1) The two-week (14-day) specification (which was the one used uniformly in the original New Mexico studies) appeared not to be long enough, in that services with the episode diagnosis that occurred more than two weeks after the initial visit were wrongly excluded from the episode; (2) relevant services occasionally occurred before the visit beginning the episode; and (3) visits, laboratory tests services, and medications were misassigned by diagnosis, leading to the irrelevant services' being included in the episode when they should be excluded, or the relevant ones' being excluded when they should be included. (The greater misassignment error was the former.)

Revisions in the New Mexico rules were then proposed that were intended to circumvent the types of "errors" seen in the Group A episodes, without introducing additional error of serious magnitude. First, the most obvious problem was that of "post-episode," services being excluded by the New Mexico rules. A reanalysis of the 20 Group A episodes in which this occurred indicated no particular pattern. Of the 20 episodes, 4 had services occurring on the 15th day, suggesting that they might have been rendered on the same day of the week two weeks later. Five episodes had services up to a week after the end of the

Table II.3

NUMBER OF GROUP A EPISODES WITH ERRORS
 WHEN ORIGINAL "NEW MEXICO" RULES
 WERE USED, BY TYPE OF ERROR

Type of Error ^{a,b}	Number of Episodes	Percent of 204 Episodes
<u>Services for Related Diagnosis</u>		
<u>Included in Episode</u>	7	3
<u>Services Missed</u>		
After end of episode	20	10
Before initial visit	8	4
<u>Services Misassigned by Diagnosis</u>		
To episode, when do not belong	11	5
Not to episode, when belong	3	1

^aCharacteristic as defined by comparing episodes created by original and revised "New Mexico" algorithms against "true" HIS episodes.

^bGroup A comprised 204 episodes--one-half of the total HIS episodes; see text for explanation.

14-day episode, and 9 episodes had services more than a week after the end of the episode. (Three episodes had missed services in two of these categories.) The first candidate for a revised rule, then, was to redefine the two-week episode as lasting 15 days (i.e., Monday through Monday), not 14 days (i.e., Monday through Sunday). No other longer length of time (e.g., three weeks defined as 22 days) appeared to be a better episode period.

The question arises as to whether a shorter episode length would be preferable. This was investigated by looking at the timing of services within two weeks among the Group A episodes. Of the 204 episodes (Table II.4), 125 (61 percent) had services rendered only on one day (i.e., the day of the index visit), and 79 (39 percent) had services on at least two different days. This number is large enough to suggest that an episode should be considered as longer than just the day of the index visit.

Of the 79 episodes with services on different days, (1) 36 had one set of services within the 14-day period in addition to those rendered on the day of the index visit; (2) 19 had two sets of additional services; and (3) 24 had services on the day of the index visit and at least one other set of services outside the 14-day period. Table II.5 shows that, among the 36 episodes with a second set of services within the 14-day period, no obvious pattern as to a predominant day emerged. The only (not unexpected) pattern is that fewer episodes had services in the second week than in the first. Of the last 24 episodes noted above in (3), 4 had also had services within the 14-day period (besides those on the initial visit) and 1 had also had two sets of additional services

Table II.4

DISTRIBUTION OF SERVICES IN ORIGINAL GROUP A EPISODES
WITHIN OR OUTSIDE THE 14-DAY EPISODE PERIOD

Days on Which Services Occurred	Number	Percent
<u>All services on day of index visit</u>	125	61
<u>Services on day of index visit</u>		
<u>and on one other day within the 14-day period</u>	36	18
<u>and on two or more other days within the 14-day period</u>	19	9
<u>and on one or more days outside the 14-day period</u>	24	12
	204	100

Table II.5

DISTRIBUTION OF SERVICES WITHIN A
14-DAY EPISODE, BY DAY OF SERVICE

Days on Which Services Occurred Within a 14-day Period	Number of Episodes
Day 2	4
3	2
Day 4	3
5	6
Day 6	6
7	1
Day 8	6
9	3
Day 10	2
11	0
Day 12	2
13	0
Day 14	1
Total^a	36

^aThese are the same 36 episodes as those in the second row of Table 4.

within the 14-day period. Given this degree of multiple services on different days, no period of time shorter than two weeks appeared to be a better episode length.

The second (relatively minor) problem was that diagnostic tests and medications were missed when they occurred before the initial visit (see Table II.3). In the original 204 Group A episodes, 6 of the 8 episodes in which this occurred involved services rendered either one or two days before the visit, suggesting that physicians might have asked patients (or parents of patients) to obtain throat cultures or to pick up a prescription and "come in the next day." Thus, a second candidate for a revised algorithm was to specify that, once an index visit for an episode was encountered, the data base be searched up to two days before the index visit for diagnostic tests and medications, and that under certain circumstances (related to diagnosis) they be included in the episode.

Diagnosis

Two additional problems involved "misassignment" by diagnosis, although they arose very infrequently, at least in the Group A episodes. The first problem involved laboratory and diagnostic tests. As explained in Chapter I, because prescription drugs never had diagnoses on the New Mexico Medicaid claim form, a specific set of rules--collectively referred to as the "prescription rule"--had been devised to assign them to episodes by diagnosis. The original New Mexico rules had also assigned laboratory tests by that same set of

steps, even if the tests had a diagnosis specified on the Medicaid claim.*

The number of HIS episodes in which laboratory and diagnostic tests were misassigned by the New Mexico rules was in fact quite small. In the original Group A, for example, only seven episodes had tests misassigned when the original New Mexico rules were used, as follows: Tests such as complete blood counts probably done to rule out infectious mononucleosis (two episodes); urinalyses for urinary tract infection or other reasons (three episodes); a chest x-ray for 'chest pain' (one episode); and a throat culture probably done for strep throat that was wrongly excluded (one episode). The number of episodes in which such tests did not have an HIS diagnosis--and would thus have needed some rule by which to be assigned--was extremely small.

Nonetheless, using a rule that potentially overrides available diagnostic information is conceptually unsatisfactory and seemingly detracts from the face validity of the episode methods. Thus, the third revision proposed was to use the diagnoses associated with the service--whenever available from the claims data--in deciding whether to include a laboratory or other test in the episode. The "prescription rule" would then be invoked only when diagnostic information was missing.

*The original decision to assign such tests by the prescription rule had been based chiefly on the observation that more laboratory tests were included in the New Mexico episodes when assigned by the prescription rules than when assigned by their own diagnoses. This may have resulted because New Mexico laboratory claim forms with no diagnoses could not be assigned to any episode and thus were lost to analysis. Using the prescription rule to assign tests was thought to give physicians the benefit of the doubt for throat cultures and other tests considered relevant to these diagnoses.

The second problem of misassignment involved the "prescription rule" itself. In the original New Mexico algorithms, prescription drugs were assigned strictly on the basis of the diagnosis of the closest preceding visit (see Chapter I); if a refill for a prescription given at the initial visit came after a visit for a totally unrelated diagnosis, the original rule did not include it in the episode. Although this specific configuration did not appear in the Group A episodes, the possibility that it might arise detracts from the general face validity of the rule. Thus, a fourth revision proposed was to modify the prescription rule such that prescriptions meeting the following criteria would be assigned to the episode: (1) All prescription drugs following a visit for either the diagnosis of the episode or a closely related diagnosis; and (2) all prescription drugs following a visit for an unrelated diagnosis if precisely the same drugs(s) had already appeared in that episode.

Group B Episodes

The Group B episodes were the mechanism by which the final set of rules for defining episodes would be tested. Reductions in the number of episodes with errors (defined as discrepancies between the HIS episodes and those created with the New Mexico algorithms [original and revised]) would lend support for modifications to the original algorithms.

Table II.6 shows that the above four revisions to the original algorithms did not reduce the number of episodes in which one or more errors were observed (see the middle two columns of the table). The

Table II.6

NUMBER OF ORIGINAL AND NEW GROUP B EPISODES WITH
 ERRORS WHEN ORIGINAL AND REVISED "NEW MEXICO"
 RULES WERE USED, BY TYPE OF ERRORS

Type of Error ^a	Original Group B				New Group B	
	Original Rules ^b		Revised Rules ^b		Revised Rules Number of Episodes	Percent
	Number of Episodes	Percent	Number of Episodes	Percent		
<u>Services for Related Diagnoses Included in Episode</u>						
	7	3	7	3	13	6
<u>Services Missed</u>						
After end of episode	22	11	22	11	2	1
Before initial visit	5	2	5	2	4	2
<u>Services Misassigned by Diagnosis</u>						
To episode, when do not belong	12	6	11	6	8	4
Not to episode, when belong	3	1	3	1	1	1

^aCharacteristic as defined by comparing episodes created by original and revised "New Mexico" algorithms against "true" HIS episodes.

^bThe original Group B comprised 204 episodes and the revised Group B comprised 212 episodes. These are the figures used to calculate percentages in this table. See text for explanation.

revised rules applied to the original Group B episodes still yielded an "error rate" of about 20 percent. The one difference was that services previously misassigned by diagnosis in one episode were now correctly assigned.

That improvement from these revisions was so modest is explained by two facts. The first is that the majority of errors were prompted by services given well after the index visit (e.g., several weeks) and were considered at this step to have been wrongly excluded from the episode. The second is that two of the revisions (these involving whether and how to use the "prescription rule") were prompted in the first place not by errors but by general conceptual and face validity concerns.

As a result, the next step was to determine the consequences of using the "clinical guidelines" developed from the physician review of the Group A episodes (described in an earlier section). The first effect was to add new but shorter episodes (by defining multiple episodes when services, especially for different diagnoses, were more than 18 days after the initial index visit); the second was to eliminate some of the original episodes. Of 13 Group B episodes to which the clinical guidelines were applicable, 1 was dropped because it was an episode of infectious mononucleosis. The remaining 12 were redefined to be 26 episodes; of the 14 "new" episodes (26 minus the original 12), 5 were not used because they had diagnoses other than one of the four sore throat diagnoses (typically acute URI). These additions and deletions yielded a final total of 212 Group B episodes.

The question still remains as to the level of error encountered when the revised New Mexico algorithms were applied to these HIS

episodes. Of the 212 new Group B episodes, 184 (87 percent) were identical when the New Mexico algorithms were used. Not surprisingly, of course, "errors" involving visits and other services missed well after the end of a two-week episode had dropped considerably, because such visits were considered to start a second (or subsequent) episode. The overall "error" rate dropped to 13 percent (28 of 212 episodes).

Of these 28 new Group B episodes in which a problem occurred, however, 13 (or 6 percent) involved services for a "closely related" respiratory infection (see the last two columns of Table II.6). These included laboratory tests or (more often) medications given for one of the following: otitis media, bronchitis, influenza, acute URI, cough, sinusitis, laryngitis and tracheitis, pneumonia, and hypertrophy of tonsils. In the remaining 15 episodes (7 percent), the most common problem (in 8 episodes, or 4 percent) was that oral medications were assigned to the episode by the prescription rule when the true reason for the prescription was ambiguous (and some question could thus be raised as to whether it belonged to the episode or not).

The frequency of the problem of misassigning prescription drugs was low; less than 5 percent of episodes had an error of this sort. Thus, exploring still more elaborate rules in an effort to overcome it did not seem to be warranted. Moreover, this was the most "common" error, suggesting that little would be gained from further manipulation or refinements of the rules. In addition, comparison of the data in Tables II.3 and II.6 suggests that, except for episodes involving services ostensibly given for closely related conditions, the numbers of errors did not rise as a result of employing these revisions and clinical guidelines.

Thus, the various modifications to the episode algorithms described in this chapter--i.e., both clinical guidelines and direct rule changes--were considered to be acceptable and useful improvements in the episode methodology. The algorithms adopted for the quality-of-care studies, which incorporate these modifications, are discussed in greater length in Chapter IV and Appendix C.

DISCUSSION

EPISODE ALGORITHMS

The validation work described in this chapter suggested that some refinement of the original episode algorithms would be well worth exploring. These refinements reduced the level of error of the original New Mexico rules without inducing appreciable "new" error. The greatest improvement came from devising "clinical guidelines" by which to shorten or clarify long and complex episodes and to eliminate those involving a process other than a simple respiratory infection.

RESTRICTING THE DATA SET

Before applying any rules to the claims data to create episodes, one "common-sense" procedure (used in all these studies) was to choose ahead of time the sets of medications, laboratory tests, and other services that would be included in the quality-of-care analyses. This was done for two practical reasons: (1) To minimize the data set that would need to be manipulated in the quality-of-care analyses, and (2) to

minimize the task of setting quality-of-care criteria that was undertaken by a panel of physicians (as described in Chapter III).

Restricting the data elements in this way is also important for reviewing insurance claims data that are likely either to have ambiguous diagnoses or to be missing diagnoses altogether, because the larger the number of elements about which an assignment mistake can be made, the higher the likelihood that a mistake will be made. In Medicaid (and Medicare) data, prescription drugs are the services most likely to lack diagnoses, because diagnoses rarely if ever appear on pharmacy claims. The larger the possible set of medications included in the analyses, therefore, the greater the probability that misassignment by diagnosis will occur, given that some algorithm must be used to assign them to episodes. This implies that only services that plausibly would be used for the conditions under study should be included in the data set.

A second reason for restricting the data set is practicality. In screening large claims files, as PSROs will do in ambulatory care review, there is little reason to incur the expense of manipulating large data files when smaller ones will provide the same level of information. For much (but not all) quality-of-care review, some services and medications may contribute little in determining whether care is acceptable or not acceptable or may be essentially irrelevant to the diagnoses under study. Some services or medications may occur infrequently for the diagnoses of interest and yet occur frequently in the full data set; two examples from the earlier New Mexico studies included tranquilizers and hormones. Without specific analytic reasons

for retaining such elements in the data file, it is more efficient not to include them for quality-of-care studies of this sort.

Deliberately editing a claims data set might be unwarranted in several situations, however. As PSROs first begin ambulatory care review, they may well need to screen all data in order to explore the capacities of the data base itself or to determine the major deficiencies in quality of care. In some circumstances, moreover, 100 percent review of all the care rendered by targeted physicians or rendered to certain types of patients may be desirable.

IMPLICATIONS OF CHOOSING AMONG EPISODE DIAGNOSES

FOR QUALITY-OF-CARE REVIEW

The problems with what diagnosis to associate with the episode itself are more troublesome. Some implications of both "progressions" and "combinations" are discussed below, especially as they relate to quality-of-care assessment in general and these New Mexico studies in particular. The implications for quality-of-care assessment of episodes in which a diagnosis changes from a less to a more serious condition are especially pertinent.

Because quality-of-care analyses depend significantly on diagnosis, one major conceptual problem with any episode methodology is how to classify and analyze episodes with two (or more) respiratory infection diagnoses or, for that matter, two (or more) diagnoses of any sort. In this discussion, "two or more diagnoses" relates to those recorded by the physician on the HIS Physician Outpatient claim for visits (and perhaps services associated with that particular visit).

The presence of multiple diagnoses may imply several things (which are not mutually exclusive): (1) Greater perceived severity of illness; (2) more than one disease process truly occurring simultaneously (e.g., strep throat and bronchitis, or otitis media and acute URI); (3) development of a true complication; (4) an attempt on the part of the diagnosing physician to be more precise in his records; (5) the presence of an acute condition superimposed on a long-term chronic condition (e.g., pharyngitis in a patient with hypertension); or (6) the presence of an acute condition occurring simultaneously with a totally unrelated acute condition (e.g., a cold and a sprain). All these possibilities (except (4)) certainly suggest that the respiratory condition may be more serious than would otherwise be the case. The theoretical and practical considerations of taking the latter two into account are important, but are beyond the scope of the quality-of-care analyses in this study.

The principal theoretical question is whether episodes with multiple diagnoses of respiratory infections--i.e., "overlapping episodes"--should be treated in any special manner. The question takes on practical importance only if the number of overlapping episodes is high and if quality-of-care judgments might differ radically depending on how the episodes were classified. The decision is basically one of choosing how to classify such an episode by diagnosis--e.g., as the "more bacterial" diagnosis, as a "complex" episode to be evaluated separately, or as two episodes (one of each diagnosis) to be evaluated separately.

Progressions

The problem of episodes of "double diagnoses" in which the diagnosis changes from one respiratory condition to another may be the easier to resolve. In the data reported in the Results section, the "strongest" pattern (if one can be said to exist in such small numbers) involved the 5 (of 421) HIS episodes of pharyngitis/tonsillitis that were "upgraded" to strep throat, in the sense that the first visit was for pharyngitis and a second or subsequent visit was for strep throat. If the quality of care of these potential "strep throat" episodes were judged by pharyngitis/tonsillitis criteria, which for this study are slightly less exacting than are the strep throat criteria,* the quality-of-care score might be biased toward a more positive judgment. Because the approach in this work has always been to give the physician the benefit of the doubt (i.e., bias results in a positive direction), this might be considered an acceptable outcome.

Disregarding the episodes of pharyngitis that changed to strep throat, a total of 12 HIS episodes (not quite 3 percent of the 421 episodes) progressed from either strep throat or pharyngitis to some other diagnosis. Of these, six changed to a condition about as serious or more so (e.g., otitis media) and six to a less serious condition. If this pattern also were replicated in the New Mexico data, the potential biases in quality-of-care scores might be offsetting, assuming they were randomly distributed across different types of physicians.

*See Chapter III. The criteria for all conditions except otitis media were less stringent than those for strep throat. Thus, any "progression" to strep throat would have this positive bias on behalf of the physician.

The pattern might not be replicated in the New Mexico data, however. There might be an even smaller proportion of episodes in which the diagnosis progressed from one respiratory disease to another, in which case the problem becomes negligible. There might, on the other hand, be a sizable number of episodes in which the diagnosis changes from a less to a more serious condition, or from one condition to another of comparable seriousness. In these cases, a true complication or exacerbation of the underlying infectious process may be occurring. This configuration of events presents problems for quality-of-care assessment if the criteria differ widely for the two conditions or if the criteria are not specific enough to deal with complications.

In these circumstances, one solution (the easier) is to eliminate from the analysis all episodes in which diagnoses progress from one condition to a more serious one (e.g., from a less to a more bacterial diagnosis, such as acute URI to pharyngitis). An alternative solution is to redefine the entire episode to be the more bacterial or more serious condition, regardless of the diagnosis of the initial visit, and to use the quality-of-care criteria relevant to the more bacterial condition. This alternative approach will be used in the quality-of-care studies despite the complexities it creates for the episode methodology, because these episodes may well be more interesting than the single diagnosis ones and, more importantly, because this approach accords more with the natural history of these infections.

(These points are taken up more fully in Chapter IV.)

Combinations

The question of "combinations" of diagnoses on claims for visits must still be addressed, because (to judge from the rate among the HIS

episodes) perhaps as many as 20 percent of all episodes may have two respiratory diagnoses specified at the index visit. The original solution to claims with pairs of diagnoses was to develop a hierarchy of common acute infectious conditions and to assign all claims with two diagnoses to the condition more likely to be bacterial in origin (as explained in Chapter I). The principal rationale for this rule was that the use of antibiotics could then be more favorably judged (leaving aside the problem of type of antibiotic). The question is whether, on empirical and theoretical grounds, this hierarchical approach is a satisfactory (or the most satisfactory) way to deal with episodes in which combinations of diagnoses appear on visit claims or whether some alternative method might be better.

Several alternative approaches are briefly discussed below. The perspective of the discussion, it should be emphasized, is one of peer review organizations such as PSROs doing quality-of-care review and screening, in which the main purpose may be detecting and correcting major deficiencies in physician performance. Thus, the constraints within which PSROs may operate--such as poorly coded claims data, few personnel trained in quality-of-care review, or lack of sophisticated computer resources--must be kept in mind.

Among the HIS sore throat episodes, combinations were much more frequent with pharyngitis/tonsillitis (somewhat more general terms for sore throat) than with strep throat (a more specific bacterial diagnosis). Moreover, the most frequent combination for pharyngitis/tonsillitis was with otitis media (again a more specific

bacterial entity). No instance of strep throat with otitis media was observed.

The New Mexico hierarchy rule would assign all combinations of diagnoses on Medicaid claims that involved strep throat to strep throat; it would then assign all remaining combinations involving otitis media to otitis media. Assigning combinations of pharyngitis/tonsillitis with otitis media to otitis media would be supported by the observation that the episode was as or more likely to be treated as otitis media than as pharyngitis. In fact, in the 14 HIS episodes in which this combination occurred, the episode was managed much more often as otitis media, as judged from the numbers of services or medications coded for otitis media. None of the combination episodes was treated solely as pharyngitis/tonsillitis, 1 was treated solely as otitis media, and of the remaining 13, 11 were treated chiefly as otitis media. Although these numbers are small (and the findings thus mainly impressionistic), they are in the direction of confirming that the present hierachial rule is a reasonable way to handle combinations.

A second approach to the problem of diagnosis combinations might be to judge such episodes by the quality-of-care criteria for each diagnosis, and to assign the episode to the diagnosis that receives the better quality score. This approach theoretically gives the benefit of the doubt to the physician and presumably is a less arbitrary approach than is the hierarchical rule. For these respiratory infections, however, it has some disadvantages.

Because strep throat and otitis media are to be judged by relatively more rigorous criteria than are the remaining diagnoses, this

second approach has the unfortunate potential for underestimating both the number of episodes of those two conditions and, more importantly, for underestimating the possible low levels of care delivered for them. For example, one criterion for high quality-of-care for otitis media is a followup visit (see Chapter III). From previous work, one can anticipate that the level of followup care will approach only about 25 percent--i.e., on average, only about 25 percent of otitis media episodes will have a return visit. If (in the "worst case") these 25 percent are all associated with another respiratory diagnosis, they would in all likelihood be assigned to that other respiratory diagnosis (because of a more favorable quality score). This distorts not only the count of episodes but also the general level of care assumed to be provided for these conditions.

Furthermore, to the degree that certain types of physicians see, on average, more or less of a given condition, the second approach also has the potential for confounding practice patterns by type of provider. To continue with the otitis media example: Assume pediatricians diagnose more otitis media than other types of physicians. A number of factors (such as predilection for more than one diagnosis, or giving generally better levels of care for all conditions) may then increase or decrease (in complex ways) the likelihood that pediatricians' quality-of-care scores will be higher, on average, than those for other physicians. Thus, the ambiguities introduced by counting episodes in one specific diagnostic category or another on the basis of better quality scores would seem to outweigh any possible benefits, and it does not appear to be a satisfactory alternative to the hierarchical rule.

A third approach might be to count episodes twice (once for each diagnosis). This has serious drawbacks, chiefly that physicians who use, e.g., contraindicated antibiotics or certain injectable drugs would be penalized twice, and physicians who use, e.g., few diagnostic tests would be rewarded (or at least not penalized) twice. For screening large data sets to determine how well providers meet even minimally acceptable quality-of-care criteria, this would be an unacceptable set of biases that could not be expected to cancel each other out.

A fourth approach is to analyze all combinations manually (i.e., on a case-by-case basis). This approach is unwieldy on two counts. First, for large data bases (even as large as the New Mexico Medicaid claims files, which by statewide Medicaid standards is not large), such a method could be prohibitively costly for peer review organizations in terms of time and personnel resources. Second, there is still no way to know what the "true" diagnosis was, meaning that even manually analyzing such episodes would require some (possibly different) set of rules by which to make decisions about final diagnosis and quality-of-care judgments.

A fifth approach would be to specify "joint" quality-of-care criteria sets for all possible (or the most commonly encountered) combinations of diagnoses. Conceptually this is the most attractive alternative, because it is apparently less arbitrary than the hierarchical rule and is a more sophisticated method from the point of view of quality assessment. This technique, however, also appears to have drawbacks.

If "joint" criteria were fairly complex and rigorous for high quality-of-care and yet essentially the same as the single-diagnosis criteria for minimal quality-of-care, scores would be biased downward to some unknown degree. If they were to require that all elements of high quality for each condition separately must also be present for quality of care to be high for the two conditions together, two problems may arise. First, the criteria may be internally contradictory (e.g., antibiotics are required for otitis media and contraindicated for acute URI). Second, criteria for high quality may be very broad and permissive (e.g., the antibiotics considered high quality for otitis media and bronchitis together cover most antibiotics except those contraindicated in all circumstances), which in turn means that minimal care becomes hard to define.

Moreover, for some combinations (e.g., pharyngitis with acute URI), joint criteria might be essentially the same as the criteria for pharyngitis alone (especially for high care), meaning that the hierarchical rule has in effect been used anyway. In addition, unless data bases were extremely large, sample sizes for "combination" episodes, although perhaps of sufficient number in the aggregate, might be too small for reliable analysis when taken combination-by-combination. This would certainly be the case when further disaggregated by physician type. Finally, much of the motivation of PSROs as they move into ambulatory care review in the next decade (and much of the purpose of the quality-of-care analyses reported in this and the earlier studies) is to determine the degree to which physicians exceed or fall below minimal levels of quality of care, and

which types of physicians exceed or fall short of those levels.

Investigating that question can be done more expeditiously when criteria specific to single diagnoses are used.

Consequently, for the purposes of the quality-of-care analyses reported in this study, using the hierarchical rule to determine the final diagnosis on claims for physician visits, laboratory tests, and the like--and hence to determine the diagnosis by which quality of care in the episode will be judged--would seem quite adequate. First, a persuasive case cannot be made that any or all of the alternatives just discussed would be better. Second, a rule of this sort could be replicated or adapted easily by organizations just beginning to review the quality of ambulatory care. Third, a rule of this sort is necessary to reduce the complexity of the data available on claims regardless of how final diagnoses for episodes are arrived at.

CHAPTER III
QUALITY-OF-CARE CRITERIA

INTRODUCTION

Three purposes motivated the work to establish quality-of-care criteria for this study. The first goal was to expand the criteria to cover all the services and medications to be included in these (more comprehensive) episodes of respiratory infection. The second objective was to modify (as necessary) the earlier criteria for antibiotic drugs. The third, and most important, motivation was define "profiles" of quality of care that would take visits, laboratory and diagnostic tests, and medications simultaneously into account.

The general approach was to solicit ratings from a number of physicians as to the level of care (from high to poor) represented by the use or nonuse of various laboratory tests and procedures, injectable and oral medications, and physician visits in these six respiratory conditions. These ratings were combined across the several physician judges to arrive at a consensus quality-of-care category for each service or medication. The various services and therapies were combined, largely on the basis of these consensus categories, into diagnosis-specific "Profiles" of High, Acceptable, Minimal, and Unacceptable care.

The remainder of this chapter notes the general considerations that guided the criteria-setting task, briefly describes the methods of

soliciting physician judgments by mailed questionnaire and computing the consensus ratings, presents the ratings and final quality-of-care Profiles, and examines briefly some implications of this part of the study. Appendix B gives more detail about the procedures used in developing these Profiles.

GENERAL CONSIDERATIONS

Several factors were considered in initiating this work. First was the recognition that criteria might differ according to the age of the patient; thus, separate criteria were solicited for infants/young children (0 to 7 years) and older children/adults (8 years and older). This split was dictated by the widely accepted contraindication of the use of tetracycline in children under 8 years of age. These are sometimes referred to below as the pediatric and adult age groups, respectively. For otitis media, separate criteria were solicited for infants/very young children (0 to 4 years) and young children (5 to 7 years) because of the known differential indications for types of antibiotics between these two age groups, derived from the differing bacterial causes of this condition. These points are covered in more detail in Lohr et al. (1980).

Second was the recognition that, conceptually, not providing a service or medication when it was warranted--an error of omission--might be considered as unsatisfactory a level of care as providing a service or medication was it was not warranted--an error of commission. The opposite is also true, of course--i.e., providing a service or drug when it is needed might be considered as satisfactory a level of care as not

providing it when it is not needed. Thus, judgments were solicited from physicians as to the level of care represented by both the presence and absence of various services/medications.

Third, "how to" set quality-of-care criteria was not explicitly addressed. For example, alternate methods of criteria-setting, such as using different questionnaires or comparing explicit versus implicit judgments, were deliberately not used. No Delphi technique was considered necessary (see, e.g., Romm and Hulka, 1979). Rather, a standardized mail-out form was developed to elicit explicit judgments about a large number of services given for these conditions from several physician raters. The form itself was pretested and modified on the basis of problems encountered initially.

Fourth, the reader will recall that the data in this study constituted virtually all ambulatory care in two six-month periods as recorded in the New Mexico Medicaid claims file. From this source, little or no information was available on either short- or long-term patient outcomes, especially outcomes that would be associated with the acute conditions studied here.* Consequently, no effort was made to develop outcome-based criteria. Instead, the emphasis of the criteria is strictly on the "process" of care. The several characteristics of physicians that have been investigated in the earlier studies (and will

*This is not to say that outcome data are always unavailable from such claims data bases or for these acute illnesses. One could, for example, imagine outcomes such as hospitalizations for pneumonia or for tonsillectomies, further ambulatory care for rheumatic heart disease, or even deaths. Looking at such outcomes (which would be very rare compared to the prevalence of the acute conditions themselves) would require a larger data base than the one available for this study and/or one that covered a longer time frame than two six-month periods.

be used in this one) form the "structural" variables of interest. In setting and using these explicit process criteria, more complex techniques such as criteria mapping or staging were not used. Such approaches are largely inapplicable to insurance claims data bases (especially when the conditions of interest are acute illnesses) because of the need for clinically relevant information usually found in medical records but beyond that usually available in claim forms.

TERMINOLOGY

Three sets of sometimes overlapping terms are used in this chapter. Because they are often used synonymously in quality-of-care work as well as in other contexts, the reader is cautioned that, for this work, they have specific meanings that are explicated later. Briefly, the quality-of-care categories in the questionnaire described below will refer to "high," "probably acceptable," "probably not acceptable," and "poor," and the median ratings described below use the same terms. In discussing these, the terms "satisfactory" and "unsatisfactory" are sometimes employed; satisfactory comprises the high/probably acceptable categories, unsatisfactory the probably not acceptable/poor. The "Profiles" referred to below are classifications of quality of care based on the median ratings derived from the questionnaire; they are designated "High," "Acceptable," "Minimal," and "Unacceptable." To distinguish the questionnaire ratings from the Profile classifications, the latter are capitalized in this chapter.

METHODSPHYSICIAN JUDGES

The judges who were asked to complete the questionnaires by which quality-of-care criteria were established were in several primary care specialties (e.g., family practice, internal medicine, pediatrics). Some were in private, office-based practice; others were in academic settings (university medical schools, Veteran Administration hospitals), often with a research component; and still others were in full-time health services research. They were from both urban and rural areas of the country.

THE QUESTIONNAIRE

Figures III.1 and III.2 reproduce (respectively) the instructions to the physician judges and the questionnaire. Six questionnaires were mailed to each judge, one questionnaire per disease; each questionnaire had separate sections for the two relevant age groups. In effect, 12 questionnaires were self-administered by each judge.

A total of 69 "variables" appeared on each questionnaire. Of these, 17 variables specified that a particular drug, test, or procedure was not given or done. Of the 69 variables, 32 referred to antibiotics, 19 to other medications, 16 to various diagnostic procedures, and 2 to physician visits. Of the 32 antibiotic variables, 1 specified no antibiotics at all, 5 were common injectable antibiotics alone, 7 were common oral antibiotics alone, and the remainder were combinations of injectable and oral antibiotics that had been (relatively) frequently observed for these conditions in the earlier

FIGURE III.1

QUALITY OF CARE FOR COMMON RESPIRATORY CONDITIONS
INSTRUCTIONS

In your judgment, what is the level of quality of care for the average patient with common respiratory infections who receives the therapies and services listed on the attached forms? The two age groups (children, adults) and the individual respiratory diagnoses are specified on each form.

There are four quality of care categories:

HIGH QUALITY OF CARE includes the following:

- using a therapy that is the best or one of the best choices to control the pathologic process or to control serious symptoms;
- using a laboratory test or other service that is usually or almost always necessary for adequate diagnosis or management of the condition;
- avoiding a therapy or service that clinically is not needed and carries a substantial risk to the patient.

PROBABLY ACCEPTABLE QUALITY OF CARE includes the following:

- using a therapy that clinically may not be required but has little or no intrinsic risk to the patient (e.g., a therapy that is potentially more beneficial than harmful or risky);
- using a laboratory test or service that is usually helpful but not absolutely necessary for adequate diagnosis or management;
- avoiding a therapy or service that clinically may not be required and carries some risk to the patient (e.g., a drug that is potentially more harmful or risky than beneficial).

PROBABLY NOT ACCEPTABLE QUALITY OF CARE includes the following:

- using a therapy or service that clinically may not be required and carries some risk to the patient (e.g., a drug that is potentially more harmful or risky than beneficial);
- using a laboratory test or service that is rarely or never necessary for adequate diagnosis or management;
- omitting a therapy or service that clinically is probably necessary for adequate diagnosis or control of pathologic processes or symptoms.

POOR QUALITY OF CARE includes the following:

- using a therapy that is contraindicated because of the high risk of serious side effects, because of age of the patient, or because of other considerations (such as improper drug combinations);
- omitting a therapy or service that clinically is necessary for adequate diagnosis or control of pathologic processes or serious symptoms.

MEDICATIONS, LABORATORY TESTS, AND SERVICES ARE LISTED IN ROWS. PLEASE PLACE ONE "X" IN EACH ROW TO INDICATE YOUR JUDGMENT AS TO THE APPROPRIATE QUALITY OF CARE CATEGORY FOR EACH ENTRY.

For example, if you believe that the average patient with strep throat who receives IM long-acting penicillin has received high quality care, place an "X" in the first column opposite the entry "IM LA Penicillin;" if you believe that the average patient with strep throat who receives IM tetracycline + oral tetracycline has received less than adequate care but not necessarily poor care, place an "X" in the third column opposite that entry. Similarly, if you believe that the average patient with strep throat who has a throat culture done has received high quality care, place an "X" in the first column opposite the entry "throat or strep culture," and if you believe that the average patient with strep throat who does not receive a throat culture has received poor quality care, place an "X" in the fourth column opposite the entry "no throat culture."

FIGURE III. 2

THIS PAGE IS BEST QUALITY PRACTICABLE
PROOF COPY FURNISHED TO DDC

episode study. Of the 19 variables involving other medications, 5 were for injectable forms. (Injectable refers only to intramuscular, or IM, drugs.)

Four quality-of-care categories were specified on the questionnaire: high, probably acceptable, probably not acceptable, and poor (see Figure III.1 for the definitions). The physician was asked to rate each service/medication as one and only one of the categories, for each disease/age group separately.

These definitions of the quality-of-care categories were designed to account for several phenomena: (1) Using a necessary or preferable service/medication (e.g., giving a correct antibiotic in strep throat); (2) not using an unnecessary or contraindicated service/medication (e.g., not using injectable analgesics); (3) not using a necessary or preferable service/medication (e.g., omitting a followup visit for otitis media); and (4) using an unnecessary or contraindicated service/medication (e.g., giving tetracycline to children). Generally speaking, (1) and (2) represent the high end of the quality-of-care spectrum, and (3) and (4) the low end of the spectrum. The definitions involving probably acceptable and probably not acceptable care were intended to cover situations in which the extremes of high or poor care could not be applied, such as using noninvasive diagnostic tests (e.g., urinalysis) or relatively innocuous drugs (e.g., oral non-narcotic analgesics). Four categories were used so that the physician judges were forced to decide, at a minimum, between satisfactory care (high,

probably acceptable) and unsatisfactory care (probably not acceptable and poor).*

Questionnaires were mailed to physicians with stamped, self-addressed envelopes. Upon being returned, the questionnaires were coded and keypunched for analysis.

MEDIAN RATINGS

Frequency distributions of the scores for each variable in each disease/age group were calculated. Based on these distributions, the simple median category for each variable was determined manually by one person and recorded. All determinations of the median were double-checked by a second recorder. In the event of discrepancies, a third (and final) check was made. Even splits between two categories (i.e., ties) were dealt with by always taking the higher rating as the median.

AGREEMENT AMONG JUDGES

The level of agreement among judges about the quality categories to which they assigned each variable in each age/diagnosis classification was studied. The proportion of agreement among all judges was calculated for five different groupings of variables for each of the 12 disease/age classifications. The five groups were the following: all 69 variables together, all 31 antibiotics (excluding no antibiotics),

*The alternative of using an odd number of categories was not adopted because it was felt that the "middle" category would be interpreted as "neither acceptable nor unacceptable" and that such an interpretation provided less information than would be available from the strategy adopted. Obviously this point could, and probably should, be tested empirically.

all 19 other medications, all 5 other medications given only in injectable form, and all 37 tests/procedures plus visits.

The proportion of agreement* was defined simply as the number of pairs of judges endorsing a specific quality-of-care rating for a specific service/medication (within a disease/age group) divided by the total number of possible agreements, aggregated across all variables or across the several subsets of variables noted above. Pairs in which one or both ratings were missing were dropped from the calculation, i.e., from both the numerator and denominator.

For these analyses, agreement was defined in two ways. First, agreement had to be "exact" between any two judges or else the pair was scored as a disagreement--i.e., both judges had to give a medication or service exactly the same rating for that particular disease/age category. Second, agreement could be "dichotomous"--i.e., merely satisfactory (high or probably acceptable) or unsatisfactory (probably not acceptable or poor). For example, a pair was scored as an agreement if one judge rated a service/medication as high and the other judge as rated it as probably acceptable. In this latter approach, a pair was scored as a disagreement only if one judge rated the service/medication either high or probably acceptable and the other judge rated it probably not acceptable or poor. The latter approach gives a measure of the general level of agreement as to the overall suitability or nonsuitability of the services/medications for these conditions, whereas

*Assuming 10 judges, for example, the proportion will be the number of agreements observed for all pairs of judges (i.e., judge 1 with judges 2 through 10, judge 2 with judges 3 through 10, and so on) divided by the total number of possible agreements (which for 69 variables and no missing data is 3105).

the former gives a more detailed measure of the level of agreement about the degree of suitability.

DIAGNOSIS-SPECIFIC PROFILES OF QUALITY OF CARE

Rating individual services/medications is essentially an atomistic approach to quality-of-care assessment, and is conceptually and heuristically less attractive than a more integrative approach. To be able to make a more holistic quality-of-care evaluation, the ratings of the various services/medications had to be combined. To this end, once the frequency distributions and median quality-of-care ratings were determined for each variable in each disease/age group, aggregate "Profiles" of High, Acceptable, Minimal, and Unacceptable care were developed for each diagnosis/age group. Several guidelines were adopted *a priori* for establishing the Profiles.

First, several factors suggested that a formal weighting scheme not be used. Little or no work in the field conclusively demonstrates what weighted scores should or might be. Payne et al. (1976) reported work on upper respiratory infection suggesting that "weights" for calculating their Physician Performance Index might range from 1 to 3 for the performance of certain procedures or use of certain services. Their work, however, provides little or no guidance as to how to weight the absence of, for example, unnecessary services or contraindicated drugs. Reidel and Reidel (1979) found simple, unweighted measures to be satisfactory for their large study of quality of ambulatory care. Further, whether conclusions about quality of care would differ meaningfully if some weighting system were used is still an open (and

Table III.1
GENERAL APPROACH FOR DEFINING QUALITY-OF-CARE PROFILES

Level of Care for Profile	Elements of Care Included In or Excluded from Profile ^a
HIGH	<p><u>Use of:</u> Specific antibiotics rated "high" (or no antibiotics at all when rated "high") Diagnostic tests or physician visits rated "high" (throat strep cultures and follow-up visits)</p> <p><u>Absence of:</u> Antibiotics rated "poor" or "probably not acceptable" Other medications rated "poor" or "probably not acceptable" Diagnostic tests rated "poor" or "probably not acceptable"</p>
ACCEPTABLE	<p><u>Use of:</u> Specific antibiotics rated "high" or "probably acceptable," (or no antibiotics at all when rated "high" or "probably acceptable") Diagnostic tests or physician visits rated "high"</p> <p><u>Absence of:</u> Antibiotics rated "poor" Other medications rated "poor"</p>
MINIMAL	<p><u>Use of:</u> Specific antibiotics rated "high", "probably acceptable," or "minimal" (or no antibiotics at all)</p>
UNACCEPTABLE	<p>Residual category if episode does not belong in one of the above categories. (The internal logic of the above categories dictates that unacceptable occurs only when antibiotics rated "poor" or "probably not acceptable" were used.)</p>

^aMedian ratings were for each disease/age category separately. These general guidelines for profiles were modified somewhat for certain conditions, as explained in the text.

empirical) question. Finally, determining empirically what physicians believe weighted scores ought to be was not pursued in this study.

Second, the same number of categories of satisfactory care (plus the category of unsatisfactory care) were desired for all diagnoses, so that comparisons across diagnoses by provider types could be made more easily.

Third, defining the range of care from High to Minimal was done as systematically as possible for all conditions. Table III.1 summarizes the overall approach. Modifications to the approach are discussed later.

Fourth, the Unacceptable category was a residual, or default, classification, in that all episodes not falling into one of the three satisfactory categories were judged Unacceptable. The internal logic of the Profiles implied that, for five of the six conditions, only episodes treated with an antibiotic rated probably not acceptable or poor was assigned to the Unacceptable category. For otitis media, initially any episode without a followup visit would also be assigned to that category.

RESULTS

JUDGES AND QUESTIONNAIRES

Eleven judges returned completed and usable questionnaires. Of these, four were in private clinical practice, four in both health services research and clinical practice, two in full-time research, and one in law school but with some clinical activity. The distribution by

specialty was as follows: three in pediatrics, two in family/general practice, one in preventive medicine, and the remainder in internal medicine. All but two were board-certified.

One judge (an internist) declined to respond to the sections of the questionnaires dealing with children ages 0 to 7 years. Thus, for all calculations, the total number of judges for the "adult" ages was 11 and for the "pediatric" ages, 10.

The 0-to-7 age group had seven disease categories because there are two age categories for otitis media; correspondingly, there are only five disease categories for the 8-and-older age group.

The 0-to-7 age group had a total of 483 items (7 diseases times 69 variables), for 4830 possible responses from the 10 judges. For this age group, 15 responses were missing: 3.1 percent of the items were missing a single response, and 0.3 percent of the responses were missing. Similarly, the 8-and-older age group had a total of 345 items (5 diseases times 69 items), for 3795 possible responses from the 11 judges. Of these, 20 responses were missing: 5.8 percent of the items were missing a single response, and 0.5 percent of the responses were missing. Thus, the missing data rate was extremely low.

Several respondents wrote qualifications or explanatory notes on their questionnaires. Two pediatricians noted that repeat throat cultures would be high or probably acceptable care in strep throat. Several physicians noted that non-narcotic analgesics would also be expected to act as antipyretics and might have been prescribed on those grounds, in which case they should be judged satisfactory care (in oral, not injectable, form). In addition, one pediatrician noted that a few

therapeutic possibilities that might be satisfactory care in otitis media had not been included on the questionnaire. One was cephalosporin in a form that was evidently not available (or very rarely used) during the period covered by this study. Another was the antibiotic combination of oral penicillin or oral erythromycin with oral sulfa. This combination seldom occurred in the previous studies, but in view of its known efficacy in the younger age group, it was given a high rating.

MEDIAN RATINGS

Tables III.2-4 show the median ratings for each service or medication for each disease/age category (Table III.2, antibiotics; Table III.3, other medications in oral and injectable form; and Table III.4, the remaining tests, procedures, and visits). These medians are taken directly from the frequency distributions of physician responses to the questionnaire, as described in the Methods section. Tables B.1 - B.12 of Appendix B give the frequency distributions on which these medians were based.

The quality-of-care criteria took account of differences by age if they arose. This proved to be less important a consideration than had been anticipated. Of all 414 "pairs" of age groups (69 variables times 6 diseases), only 25 (6 percent) showed any difference in median ratings by age group. Of these 25, 15 involved antibiotics, 3 other medications, and 7 various diagnostic procedures and physician visits. Ten of the 25 differences by age were observed for otitis media (all antibiotics), 7 for bronchitis, 3 each for strep throat and influenza, and 2 for pharyngitis. Fourteen differences by age group (3 percent)

Table III.2a

MEDIAN QUALITY OF CARE RATINGS FOR SERVICES AND MEDICATIONS
USED IN COMMON RESPIRATORY INFECTIONS, BY TYPE OF
ANTIBIOTIC, DIAGNOSIS, AND AGE

Type of Antibiotic	Median Quality of Care Ratings ^{a,b}					
	Strep Throat		Otitis Media		Pharyngitis	
	0-7	8+	0-4	5-7	0-7	8+
No antibiotics	Poor	Poor	Poor	Poor	PA	PA
IM SA Penicillin	Poor	Poor	Poor	PNA	Poor	Poor
IM LA Penicillin	High	High	Poor	PA	PA	PA
IM Ampicillin	PNA	PNA	PA	PNA	PNA	PNA
IM Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
IM Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Penicillin	High	High	PNA	High	PA	PA
Oral Ampicillin	PA	PNA	High	High	PNA	PNA
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Erythromycin	High	High	PA	PA	PA	PA
Oral Cephalosporin	PNA	PNA	Poor	Poor	PNA	PNA
Oral Sulfa	Poor	Poor	PNA	PNA	Poor	Poor
IM SA Penicillin +						
Oral Penicillin	PA	PA	PNA	High	PA	PA
Oral Ampicillin	PNA	PNA	PA	PA	PNA	PNA
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Erythromycin	PA	PA	PNA	PA	PA	PA
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
IM LA Penicillin +						
Oral Penicillin	PA	PA	PNA	PA	PNA	PNA
Oral Ampicillin	PNA	PNA	PA	PA	PNA	PNA
Oral Erythromycin	PNA	PNA	PNA	PA	PNA	PNA
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Sulfa	Poor	Poor	PNA	Poor	Poor	Poor
IM Ampicillin +						
Oral Penicillin	PNA	PNA	PA	PA	PNA	PNA
Oral Ampicillin	PNA	PNA	High	PA	PNA	PNA
IM Lincomycin +						
Oral Penicillin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Ampicillin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Erythromycin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
IM Tetracycline +						
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable.

^bThe total number of judges for the pediatric (0-7 years) age group was 10; the total for the 8 years and older group was 11. In the event of "ties," the higher rating was used. See Appendix A for detail.

Table III.2b

MEDIAN QUALITY OF CARE RATINGS FOR SERVICES AND MEDICATIONS
USED IN COMMON RESPIRATORY INFECTIONS, BY TYPE OF
ANTIBIOTIC, DIAGNOSIS, AND AGE

Type of Antibiotic	Median Quality of Care Ratings ^{a,b}					
	Bronchitis		Influenza		Acute URT	
	0-7	8+	0-7	8+	0-7	8+
No antibiotics	PA	PA	High	High	High	High
IM SA Penicillin	PNA	PNA	Poor	Poor	Poor	Poor
IM LA Penicillin	PNA	PNA	Poor	Poor	Poor	Poor
IM Ampicillin	PNA	PNA	Poor	Poor	Poor	Poor
IM Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
IM Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Penicillin	PA	PA	PNA	PNA	Poor	Poor
Oral Ampicillin	PA	PA	PNA	Poor	Poor	Poor
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Tetracycline	Poor	PA	Poor	Poor	Poor	Poor
Oral Erythromycin	PNA	PA	Poor	Poor	Poor	Poor
Oral Cephalosporin	Poor	PNA	Poor	Poor	Poor	Poor
Oral Sulfa	Poor	Poor	Poor	Poor	Poor	Poor
IM SA Penicillin +						
Oral Penicillin	PNA	PNA	Poor	Poor	Poor	Poor
Oral Ampicillin	PNA	PNA	Poor	Poor	Poor	Poor
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Erythromycin	PNA	PNA	Poor	Poor	Poor	Poor
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
IM LA Penicillin +						
Oral Penicillin	PNA	PNA	Poor	Poor	Poor	Poor
Oral Ampicillin	PNA	PNA	Poor	Poor	Poor	Poor
Oral Erythromycin	PNA	PNA	Poor	Poor	Poor	Poor
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Sulfa	Poor	Poor	Poor	Poor	Poor	Poor
IM Ampicillin +						
Oral Penicillin	PNA	Poor	Poor	Poor	Poor	Poor
Oral Ampicillin	PA	PA	Poor	Poor	Poor	Poor
IM Lincomycin +						
Oral Penicillin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Ampicillin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor
Oral Erythromycin	Poor	Poor	Poor	Poor	Poor	Poor
Oral Lincomycin	Poor	Poor	Poor	Poor	Poor	Poor
IM Tetracycline +						
Oral Tetracycline	Poor	Poor	Poor	Poor	Poor	Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable.

^bThe total number of judges for the pediatric (0-7 years) age group was 10; the total for the 8 years and older group was 11. In the event of "ties," the higher rating was used. See Appendix A for detail.

Table III.3a

MEDIAN QUALITY OF CARE RATINGS FOR SERVICES AND MEDICATIONS
USED IN COMMON RESPIRATORY INFECTIONS, BY TYPE OF
MEDICATION, DIAGNOSIS, AND AGE

Type of Other Medications	Median Quality of Care Ratings ^{a,b}					
	Strep Throat		Otitis Media		Pharyngitis	
	0-7	8+	0-4	5-7	0-7	8+
IM Antitussives	Poor	Poor	Poor	Poor	Poor	Poor
Oral Antitussives	PA	PA	PA	PA	PA	PA
No Antitussives	High	High	High	High	High	High
IM Antihistamines	Poor	Poor	Poor	Poor	Poor	Poor
Oral Antihistamines	PNA	PNA	PA	PA	PNA	PNA
No Antihistamines	High	High	PA	PA	High	High
IM Antinauseants	Poor	Poor	Poor	Poor	Poor	Poor
Oral Antinauseants	Poor	PNA	PA	PA	PNA	PNA
No Antinauseants	High	High	High	High	High	High
IM Narcotic Analgesics	Poor	Poor	Poor	Poor	Poor	Poor
Oral Narcotic Analgesics	Poor	Poor	PNA	PNA	Poor	Poor
No Narcotic Analgesics	High	High	High	High	High	High
IM Non-narcotic Analgesics	Poor	Poor	Poor	Poor	Poor	Poor
Oral Non-narcotic Analgesics	PA	PA	High	High	PA	PA
No Non-narcotic Analgesics	High	High	PA	PA	PA	PA
Oral Decongestants	PA	PA	PA	PA	PA	PA
No Decongestants	High	High	PA	PA	High	High
Bronchial Dilators	PNA	Poor	Poor	Poor	Poor	Poor
No Bronchial Dilators	High	High	High	High	High	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric (0-7 years) age group was 10; the total for the 8 years and older group was 11. In the event of "ties," the higher rating was used. See Appendix A for detail.

AD-A103 341

RAND CORP SANTA MONICA CA
QUALITY OF CARE IN EPISODES OF COMMON RESPIRATORY INFECTIONS IN--ETC(U)
OCT 80 K N LOHR
UNCLASSIFIED RAND/P-6570

F/G 6/5

NL

2 of 4
4D A
108341

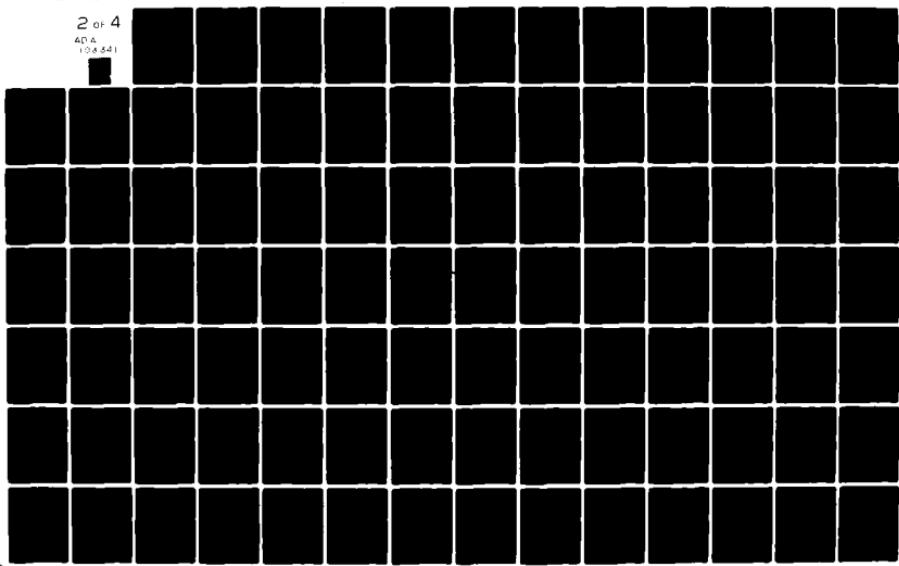


Table III.3b

MEDIAN QUALITY OF CARE RATINGS FOR SERVICES AND MEDICATIONS
USED IN COMMON RESPIRATORY INFECTIONS, BY TYPE OF
MEDICATION, DIAGNOSIS, AND AGE

Type of Other Medications	Median Quality of Care Ratings ^{a,b}					
	Bronchitis		Influenza		Acute URI	
	0-7	8+	0-7	8+	0-7	8+
IM Antitussives	Poor	Poor	Poor	Poor	Poor	Poor
Oral Antitussives	PA	PA	PA	PA	PA	PA
No Antitussives	PA	PA	PA	PA	High	High
IM Antihistamines	Poor	Poor	Poor	Poor	Poor	Poor
Oral Antihistamines	PNA	PNA	PA	PA	PA	PA
No Antihistamines	PA	PA	High	High	High	High
IM Antinauseants	Poor	Poor	Poor	Poor	Poor	Poor
Oral Antinauseants	PNA	PNA	PA	PNA	PNA	PNA
No Antinauseants	PA	PA	PA	PA	PA	PA
IM Narcotic Analgesics	Poor	Poor	Poor	Poor	Poor	Poor
Oral Narcotic Analgesics	Poor	Poor	Poor	Poor	Poor	Poor
No Narcotic Analgesics	High	High	High	High	High	High
IM Non-narcotic Analgesics	Poor	Poor	Poor	Poor	Poor	Poor
Oral Non-narcotic Analgesics	PA	PA	PA	PA	PA	PA
No Non-narcotic Analgesics	PA	PA	PA	PA	PA	PA
Oral Decongestants	PA	PA	PA	PA	PA	PA
No Decongestants	PA	PA	PA	PA	PA	PA
Bronchial Dilators	PA	PA	PNA	PNA	PNA	PNA
No Bronchial Dilators	PA	PA	PA	PA	High	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric (0-7 years) age group was 10; the total for the 8+ years and older group was 11. In the event of "ties," the higher rating was used. See Appendix A for detail.

Table III.4a

MEDIAN QUALITY OF CARE RATINGS FOR SERVICES AND MEDICATIONS
USED IN COMMON RESPIRATORY INFECTIONS, BY TYPE OF
SERVICE, DIAGNOSIS, AND AGE

Type of Service	Median Quality of Care Ratings ^{a,b}					
	Strep Throat		Otitis Media		Pharyngitis	
	0-7	8+	0-4	5-7	0-7	8+
Chest X-ray	PNA	PNA	PNA	PNA	PNA	PNA
No Chest X-ray	High	High	High	High	High	High
Sinus X-ray	PNA	PNA	PNA	PNA	PNA	PNA
No Sinus X-ray	High	High	High	High	High	High
WBC/CBC	PA	PA	PA	PA	PA	PNA
No WBC/CBC	High	High	PA	PA	High	High
Heterophile	PA	PA	PNA	PNA	PA	PNA
No Heterophile	High	High	PA	PA	High	High
Urinalysis	PNA	PNA	PNA	PNA	PNA	PNA
No Urinalysis	High	High	PA	PA	High	High
Strep/Throat Culture	High	High	PA	PA	High	High
No Strep/Throat Culture	PNA	PNA	PA	PA	PA	PA
Other Culture	PNA	PNA	PA	PA	PNA	PNA
No Other Culture	High	High	PA	PA	High	High
Panel/Profile	PNA	PNA	PNA	PNA	PNA	PNA
No Panel/Profile	High	High	High	High	High	High
Followup Visits	PA	PA	High	High	PA	PA
No Followup Visits	PA	PA	Poor	Poor	PA	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric (0-7 years) age group was 10; the total for the 8 years and older group was 11. In the event of "ties," the higher rating was used. See Appendix A for detail.

Table III.4b

MEDIAN QUALITY OF CARE RATINGS FOR SERVICES AND MEDICATIONS
USED IN COMMON RESPIRATORY INFECTIONS, BY TYPE OF
SERVICE, DIAGNOSIS, AND AGE

Type of Service	Median Quality of Care Ratings ^{a,b}					
	Bronchitis		Influenza		Acute URI	
	0-7	8+	0-7	8+	0-7	8+
Chest X-ray	PA	PNA	PA	PA	PNA	PNA
No Chest X-ray	PA	PA	PA	PA	High	High
Sinus X-ray	PNA	PNA	PNA	PNA	PNA	PNA
No Sinus X-ray	PA	PA	PA	PA	High	High
WBC/CBC	PA	PA	PA	PA	PNA	PNA
No WBC/CBC	PA	PA	PA	PA	PA	High
Heterophile	PNA	PNA	PNA	PNA	PNA	PNA
No Heterophile	PA	PA	PA	PA	High	High
Urinalysis	PNA	PNA	PNA	PNA	PNA	PNA
No Urinalysis	High	High	PA	PA	High	High
Strep/Throat Culture	PA	PA	PA	PA	PA	PA
No Strep/Throat Culture	High	High	PA	PA	High	High
Other Culture	PA	PA	PNA	PNA	PNA	PNA
No Other Culture	PA	PA	PA	PA	PA	PA
Panel/Profile	PNA	PNA	PNA	PNA	PNA	PNA
No Panel/Profile	PA	High	PA	High	High	High
Followup Visits	High	PA	PA	PA	PNA	PNA
No Followup Visits	PNA	PA	PA	PA	PA	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric (0-7 years) age group was 10; the total for the 8 years and older group was 11. In the event of "ties," the higher rating was used. See Appendix A for detail.

involved a split between satisfactory and unsatisfactory care. Most of these were observed for the antibiotics in otitis media (see Table III.2a). Given the epidemiologic evidence that different bacteria are likely to cause otitis media in younger or older children, these differences are not surprising.

The median ratings conveyed more information for some services and medications than for others, and for certain disease/age combinations than for others. This was true especially when the ratings for using the service/medication were considered simultaneously with the ratings for not using the same service/medication. For example, no mistake can be made about the consensus that antibiotics in acute URI and influenza are unsatisfactory care: as can be seen in the last four columns of Table III.2b, not using them was rated high and all the individual antibiotics were rated poor or (in a few cases) probably not acceptable. The role of different antibiotics for otitis media, by contrast, is much more mixed (see the middle two columns of Table III.2a). As another example, the median ratings (poor) about the use of other medications in injectable form (see the relevant rows of Tables III.3a and III.3b) leave no doubt as to their unacceptability for all these conditions. In contrast, the median ratings for drugs such as oral non-narcotic analgesics and decongestants in all conditions indicated that their use or nonuse is essentially irrelevant to any evaluation of the quality of care, because both use and nonuse were typically rated probably acceptable (again, see Tables III.3a and III.3b). The same phenomenon--probably acceptable ratings for both use and nonuse of certain services, leading to the conclusion that they are

less critical for quality-of-care assessment--was observed for some diagnostic tests and visits. Examples (see Tables III.4a and III.4b) include white or complete blood counts in all conditions except acute URI and for followup visits for all conditions except otitis media and bronchitis in children 0 to 7 years of age.

AGREEMENT AMONG JUDGES

To gain some appreciation of the level of agreement among the judges, the proportion of agreement was calculated for five categories of variables: all 69 services and medications together, 31 antibiotics, 19 other medications, 5 other medications in injectable form only, and 37 tests/procedures and followup visits.

As noted in the Methods section, two ways of defining agreement were used. The first was "exact," in the sense that to have 100 percent agreement, all judges had to agree on the precise rating for each service/drug under consideration. This is a stringent definition of agreement, because only slightly different opinions among the judges--generated perhaps simply by tendencies to be conservative or not conservative in treatment or use of diagnostic tests--can produce a great many disagreements, and thus a relatively meager level of exact agreement. "Random" error in the sense of isolated instances of irrational or careless responses will also lower the level of exact agreement.

The second definition of agreement was "dichotomous," in the sense that to have 100 percent agreement, all judges had to agree only that the service or medication represented either satisfactory (high and/or

Table III.5

PERCENT AGREEMENT AMONG JUDGES AS TO QUALITY OF CARE
RATINGS FOR SETS OF SERVICES AND/OR MEDICATIONS

Disease/Age Category	All Services and Medications		All Antibiotics		All Other Medications		All Tests and Followup Visits		All Injectable Other Medications	
	Exact*	Dichot.**	Exact	Dichot.	Exact	Dichot.	Exact	Dichot.	Exact	Dichot.
<u>Strep Throat</u>										
0-7 years	52	79	54	81	52	78	47	77	82	93
8+ years	49	77	49	79	51	74	45	75	81	92
<u>Pharyngitis or Tonsillitis</u>										
0-7 years	51	78	53	77	57	81	45	78	92	100
8+ years	47	76	46	75	52	77	45	79	80	96
<u>Otitis Media</u>										
0-4 years	52	80	53	75	53	83	48	85	77	92
5-7 years	50	79	50	74	53	83	48	84	73	92
<u>Acute Bronchitis</u>										
0-7 years	50	77	53	77	53	76	43	76	74	81
8+ years	44	73	44	69	47	77	42	76	59	80
<u>Influenza</u>										
0-7 years	56	87	62	99	55	76	47	76	74	81
8+ years	48	82	49	87	49	77	45	80	61	80
<u>Acute URI</u>										
0-7 years	58	89	70	98	51	83	42	78	70	100
8+ years	50	85	55	88	47	82	44	81	58	89

* Exact means that each pair of judges rated the services and/or medications exactly the same way for each age/disease category.

** Dichotomous means that each pair of judges rated the services and/or medications as either "high or probably acceptable" or "probably not acceptable or "poor" quality of care, but may have disagreed within either level.

probably acceptable) or unsatisfactory (probably not acceptable and/or lpoor) care. This is a much less rigorous definition of agreement than that just discussed.

Table III.5 gives the proportion of both exact and dichotomous agreement for all variables taken together and for the four subsets of variables. (Recall that the 0-to-7 age group refers to 10 judges, and the 8-and-older group to 11 judges.) Taking all 69 services together, the proportion of exact agreement for the 0-to-7 age group ranged from 50 percent (otitis media, 5-to-7 years, and bronchitis) to 58 percent (acute URI); for the 8-and-older group, the range was 44 percent (bronchitis) to 50 percent (acute URI).

Exact agreement was higher for injectable forms of symptomatic medications (column 9 of Table III.5) than for any other subset (columns 3, 5, and 7). For injectable symptomatic drugs, exact agreement ranged from 70 to 92 percent for the 0-to-7 age group and 58 to 81 percent for the 8-and-older group. Agreement tended to be higher for medications (in any form) than for tests/procedures/visits, but only trivially so in some cases. The proportion of exact agreement about antibiotics ranged from 53 percent (several conditions) to 70 percent (acute URI) for the 0-to-7 age group, and from 44 (bronchitis) to 55 percent (acute URI) for the 8-and-older age group.

*Column 5 includes both oral and injectable forms of these medications. If the agreements for the injectable forms were deleted from the calculation for this column, the remaining figures for the oral forms only would be lower than the figures shown for both forms together.

The percentages of "dichotomous" agreement shown in Table III.5 were higher because disagreements between contiguous quality ratings (except for probably acceptable and probably not acceptable) were not counted as disagreements. Overall, agreement ranged from 77 to 89 percent for the 0-to-7 age group and from 73 to 85 percent for the 8-and-older age group. The same general patterns noted above with respect to agreement about subsets of services/medications were of course also observed.

These levels of agreement among judges were sufficiently high that a good deal of confidence could be placed in the median consensus ratings to be used in defining the disease/age-specific Profiles of quality of care. This conclusion was based on three considerations. First, the figures given in Table III.5 for exact agreement corresponded roughly to a pattern in which 7 of the 10 (or 11) judges agreed as to the precise quality rating. In setting quality criteria, that pattern is often taken as acceptable evidence of agreement on any given criterion. The figures for dichotomous agreement exceeded that standard, in some cases substantially so. Second, although "typical" physicians did not participate in the criteria-setting task, the physician judges (who draw on strong academic backgrounds) nevertheless currently practice in widely differing settings and for quite different patient populations.

Third, information aggregated across several judges is more reliable than information from any one judge. (This is the conceptual equivalent of saying that reliability of a scale or test is increased as the number of items is increased.) To the degree that any "true" consensus might exist, an "average" over several judges is more likely

to be closer to it than is the response of any single judge. Moreover, variables with lower agreement for any given diagnosis may have been too broad or nonspecific (e.g., non-narcotic analgesics; followup visits), but in these instances the median was likely to be probably acceptable or probably not acceptable, in which case the variables did not figure importantly in the Profiles.

QUALITY-OF-CARE PROFILES

Tables III.6-III.11 give the quality-of-care Profiles for each disease. Three levels of generally satisfactory care--High, Acceptable, and Minimal--were defined for each condition, and Unacceptable was the residual category.* The reader is referred to Table III.1 for the general approach taken in defining these levels of care.

As mentioned above, within diseases the differences in criteria by age were relatively few, and several were differences only in degree of acceptability or nonacceptability. Profiles for acute URI, influenza, and pharyngitis were not differentiated by age group. Profiles for the other three diagnoses differed by age only as indicated in Tables III.6, III.7, and III.9.

"No antibiotics" had been rated poor care for strep throat.** This was considered too rigid a criterion. Some providers may use the strep

*If a given element of care was rated as any level of appropriate care, it was automatically allowed at all lower levels of appropriate care. For instance, any antibiotic included as part of High quality was also Acceptable and Minimal.

**The physician questionnaire did not contain entries for "no antibiotics with culture" and "no antibiotics without culture."

throat diagnosis as a rule-out, await the results of a throat or strep culture before deciding on therapy, and in the event of a negative culture, quite properly not use antibiotics. Such a practice pattern would be quite good care. Thus, the Profiles for strep throat specify that quality-of-care is High if a throat or strep culture is done and no antibiotics are given (and all other requirements for High care are met).

The same reasoning was applied to pharyngitis. Because this condition is less likely to be bacterial than is strep throat, there is some justification for being less rigorous about not using antibiotics. Therefore, "no antibiotics" even without a throat or strep culture was classified as Minimal care.

In a few cases, an antibiotic initially rated Unacceptable care was upgraded to at least Minimal care (or higher). Two examples follow. In strep throat, IM short-acting (SA) penicillin plus oral erythromycin was rated probably not acceptable. The situation could arise, however, in which this combination was observed precisely because the patient had an allergic reaction to the injectable penicillin and then subsequently (and properly) received a suitable substitute for the penicillin (namely, erythromycin). Because it is difficult to know from claims data whether this set of circumstances explains the occurrence of this combination, it was allowed as Minimal care. For pharyngitis, IM long-acting (LA) penicillin plus oral penicillin had been rated probably not acceptable, whereas IM SA penicillin plus oral penicillin had been rated probably acceptable. Although the former combination represents somewhat redundant therapy (because one injection of LA penicillin is

Table III.6
BASIC PROFILE FOR STREPTOCOCCAL SORE THROAT: ALL AGES

HIGH	ACCEPTABLE ^a	MINIMAL ^a
IM LA penicillin Oral penicillin Oral erythromycin IM SA penicillin with oral penicillin No antibiotics at all	or or or or or No antibiotics at all	IM LA penicillin alone, or with oral penicillin or IM SA penicillin with oral penicillin or Oral penicillin or Oral erythromycin or Oral penicillin or Oral erythromycin
PLUS	PLUS	PLUS
Strep or throat culture	Strep or throat culture or None of the following: Oral antihistamines, antinauseants, narcotic analgesics, bronchial dilators	None of the following: Oral antinauseants, narcotic analgesics IM antitussives, antihistamines, antinauseants, narcotic analgesics non-narcotic analgesics
		Chest or sinus x-ray, urinalysis/urine culture other cultures panel/profile tests

^a For children ages 0 to 7 years, Acceptable and Minimal also included oral ampicillin.

Table III.7a
BASIC PROFILE FOR OTTIS MEDIA
AGES 0 TO 4 YEARS

ACCEPTABLE		MINIMAL	
HIGH			
oral ampicillin with or without oral sulfa	Oral ampicillin with or without oral sulfa	Oral ampicillin with or without oral sulfa	Oral ampicillin with or without oral sulfa
oral penicillin with oral sulfa	Oral penicillin with oral sulfa	Oral penicillin with oral sulfa	Oral penicillin with oral sulfa
oral erythromycin with oral sulfa	Oral erythromycin with oral sulfa	Oral erythromycin with oral sulfa	Oral erythromycin with oral sulfa
IM ampicillin with oral ampicillin	IM ampicillin with oral ampicillin	IM ampicillin with oral ampicillin	IM ampicillin with oral ampicillin
<u>PLUS</u>			
Followup visit			
IM antitussives, antihistamines, antinauseants, narcotic analgesics, non-narcotic analgesics	<u>None of the following:</u> <u>Oral narcotic analgesics,</u> <u>bronchial dilators</u>	<u>None of the following:</u> <u>Oral bronchial dilators</u>	
Chest or sinus x-ray, heterophile, urinalysis/urine culture, panel/profile tests			
	<u>PLUS</u>	<u>PLUS</u>	<u>PLUS</u>
	Followup visit	Followup visit	Followup visit

Table III.7b

BASIC PROFILE FOR OTITIS MEDIA
AGES 5 TO 7 YEARS

HIGH	ACCEPTABLE	MINIMAL
Oral penicillin with or without oral sulfa or Oral ampicillin with or without oral sulfa or IM SA penicillin with oral penicillin	Oral penicillin with or without oral sulfa or Oral ampicillin with or without oral sulfa or Oral erythromycin with or without oral sulfa or IM LA penicillin alone, or with oral penicillin, oral ampicillin, or oral erythromycin	Oral penicillin with or without oral sulfa or Oral ampicillin with or without oral sulfa or Oral erythromycin with or without oral sulfa or IM LA penicillin alone, or with oral penicillin, oral ampicillin, or oral erythromycin
<u>PLUS</u>		
Followup visit		
<u>PLUS</u>		
None of the following: Oral narcotic analgesics, bronchial dilators	IM SA penicillin with oral penicillin, oral ampicillin, or oral erythromycin	IM SA penicillin with oral penicillin, oral ampicillin, or oral erythromycin
IM antitussives, antihistamines, antinauseants, narcotic analgesics	IM ampicillin with oral penicillin or oral ampicillin	IM ampicillin with oral penicillin or oral ampicillin
non-narcotic analgesics	<u>PLUS</u>	<u>PLUS</u>
Chest or sinus x-ray, heterophile urinalysis/urine culture panel/profile tests	Followup visit	Followup visit
	<u>PLUS</u>	<u>PLUS</u>
	None of the following: Oral bronchial dilators	IM antitussives, antihistamines, antinauseants, narcotic analgesics non-narcotic analgesics

Table III.8

BASIC PROFILE FOR PHARYNGITIS/TONSILLITIS
A.I.J. AGES

HIGH	ACCEPTABLE	MINIMAL
IM LA penicillin Oral penicillin Oral erythromycin IM SA penicillin with oral penicillin No antibiotics at all	or IM LA penicillin alone, with oral penicillin Oral penicillin Oral erythromycin IM SA penicillin with oral penicillin No antibiotics at all	or IM LA penicillin alone, with oral penicillin Oral penicillin Oral erythromycin IM SA penicillin with oral penicillin or oral erythromycin No antibiotics at all
<u>PLUS</u>		
Strep or throat culture	Strep or throat culture	
<u>PLUS</u>		
None of the following: Oral antihistamines, antinauseants, narcotic analgesics, bronchial dilators	None of the following: Oral narcotic analgesics, bronchial dilators	
IM antitussives, antihistamines, antinauseants, narcotic analgesics, non-narcotic analgesics	IM antitussives, antihistamines, antinauseants, narcotic analgesics non-narcotic analgesics	
Chest or sinus x-ray, urinalysis/urine culture, other cultures, panel/profile tests		

Table III.9
BASIC PROFILE FOR BRONCHITIS: ALL AGES

HIGH ^a	ACCEPTABLE ^a	MINIMAL
No antibiotics Oral ampicillin <u>PLUS</u> Followup visit ^b	or No antibiotics Oral ampicillin with or without IM ampicillin Oral Penicillin Oral erythromycin	or No antibiotics Oral ampicillin with or without IM ampicillin Oral penicillin Oral erythromycin
<u>PLUS</u>	<u>PLUS</u>	
<u>None of the following:</u> <u>Oral</u> antihistamines, antinauseants, narcotic analgesics	Followup visit ^b	
IM antitussives, antihistamines, antinauseants, narcotic analgesics non-narcotic analgesics	or None of the following: <u>Oral</u> narcotic analgesics	IM antitussives, antihistamines, antinauseants, narcotic analgesics non-narcotic analgesics
Stim x-ray, heterophile, urinalysis/urine culture panel/profile tests		

^aFor persons 8 years and older, High also included oral penicillin, oral erythromycin, and oral tetracycline, and Acceptable and Minimal also included oral tetracycline.

^bFollowup visits were required only for children 0 to 7 years of age.

Table III.10
BASIC PROFILE FOR INFLUENZA
ALL AGES

HIGH	ACCEPTABLE	MINIMAL
<p>No antibiotics</p> <p>PLUS</p> <p>None of the following:</p> <p>Oral narcotic analgesics, bronchial dilators,</p> <p>IM antitussives, antihistamines, antinauseants, narcotic analgesics, non-narcotic analgesics</p> <p>Sinus x-ray, heterophile, urinalysis/urine culture, other cultures, panel/profile tests</p>	<p>No antibiotics</p> <p>PLUS</p> <p>None of the following:</p> <p>Oral narcotic analgesics</p> <p>IM antitussives, antihistamines, antinauseants, narcotic analgesics, non-narcotic analgesics</p>	<p>No antibiotics</p>

Table III.11
BASIC PROFILE FOR ACUTE URI
ALL ACTS

HIGH	ACCEPTABLE	MINIMAL
No antibiotics <u>PLUS</u> None of the following: <u>Oral</u> antinauseants, narcotic analgesics, bronchial dilators <u>IM</u> antitussives, antihistamines, antinauseants, narcotic analgesics, non-narcotic analgesics Chest or sinus x-ray, heterophile, urinalysis/urine culture, other cultures, panel/profile tests	No antibiotics <u>PLUS</u> None of the following: <u>Oral</u> narcotic analgesics IM antitussives, antihistamines, antinauseants, narcotic analgesics, non-narcotic analgesics	No antibiotics

sufficient), it was upgraded to Acceptable on two grounds: (1) Nothing is egregious about the therapy (e.g., it involves neither two types of antibiotics nor a contraindicated antibiotic) and (2) upgrading avoids penalizing a provider for a data processing error in which SA and LA injectable penicillin were confused.

Some antibiotics that had a median rating of only probably acceptable were upgraded to High quality in the Profiles for bronchitis and pharyngitis. The upgrading was done for two reasons: (1) To be able to define a High quality-of-care Profile more fully than would otherwise have been possible, and (2) to take account of differences in the degree to which certain antibiotics were rated probably acceptable. For pharyngitis, as one example, IM SA pencillin plus either oral penicillin or oral erythromycin had median ratings of probably acceptable. The former combination, however, had only two individual ratings of probably not acceptable or poor, whereas the latter had five such individual ratings. Thus, the former combination was allowed as High quality in the Profile, and the latter was allowed only as Minimal.

DISCUSSION

These Profiles were used to examine differences in quality of care by provider type and over time, as described in the following chapter, with an emphasis on levels of satisfactory care, not levels of deficient care. The sets of criteria that defined Acceptable Profiles allowed a fair amount of flexibility to the physician, for example by not penalizing him or her for any diagnostic work that might be done, no

matter how peripheral (or costly) it might seem. As a general guide, about all that the Acceptable Profiles disallowed were antibiotics that are distinctly unsuitable for these types of infections and various injectable drugs--and both of these exclusions would be consistent with the original EMCRO guidelines. Followup visits for otitis media and bronchitis (for children) were emphasized, as were throat cultures for strep throat and pharyngitis/tonsillitis. The Minimal criteria care were related virtually to only the use of contraindicated antibiotics. Providers whose care falls into this particular category may in fact be practicing quite marginal medicine.

These Profiles had two disadvantages. The first was that the criteria for Minimal care were, in fact, fairly minimal. Apart from specifying that only acceptable or high antibiotics (or no antibiotics) must be used (and probably not acceptable or poor antibiotics avoided), they permitted a variety of dubious practices (for example, giving medications such as antitussives or analgesics in injectable form). The second was that they do not take account of possible differences in degree of quality of care represented by situations in which several poor elements of care occur versus situations in which only one poor element occurs. As a case in point, one episode may receive an antibiotic rated poor care and two other medications both rated poor care, whereas another episode may receive simply the poor antibiotic. Another case (which for this population may be observed in well over 50 percent of the otitis media episodes) is that some episodes may be rated Unacceptable because of both no followup visit and poor antibiotics, whereas others may have only one of the problems. In some sense, the

first examples in these two cases represent "worse" care than the latter, but the Profiles by themselves do not detect the distinction.

These disadvantages of the Profiles were considered tolerable here for several reasons. First, defining more quality-of-care levels for these diseases--i.e., still more complex configurations of services/medications being present and/or absent--would result in a proliferation of quality-of-care categories that could be less easily compared across diagnoses and provider types and that might have sample sizes too small for reliable analysis. Limiting the number of classifications in the Profiles was preferred for ease of presentation and understanding and for practical analytic considerations. (As will be seen in Chapter IV, even three levels of adequate care proved unnecessary for this study. Virtually all episodes of conditions other than otitis media were characterized as High, Acceptable, or Unacceptable.)

Second, the emphasis of this study was positive, not negative. Attention was directed to (1) whether over time and among provider types, quality of care changed from Unacceptable to at least Minimally acceptable, and (2) whether, and from whom, exemplary care was delivered. Investigating whether basically unsatisfactory care from certain provider categories changes, e.g., from "very poor" to "less poor but still unacceptable" would not seem to be a very interesting or productive exercise, at least not in comparison with the resources that would be needed to explore the question adequately.

Cost is one health policy issue that these criteria do not take into account. Some thought was given to asking the physician judges to

take the monetary costs of these services/medications into consideration in their rating task. For example, they could have been asked to factor into their decision the relative expensiveness or inexpensiveness of different antibiotics as they selected one of the four quality categories on the questionnaire.* This finally was not done, for several reasons.

First, it would have complicated and lengthened the rating task for the judges. Second, for the purposes of this study (applying the criteria to the care given by physicians serving a Medicaid population in the early 1970s), it was felt that "cost-conscious" criteria might not be entirely applicable, because of the possibility that such physicians, at least then, did not take cost to the Medicaid patient (or program) into consideration in their diagnostic or therapeutic choices. Third, depending on the circumstances in which the physician judges practiced, their experience with relative (or at least absolute) prices for drugs and/or diagnostic tests might differ; had they taken cost overtly into account, this factor might, in turn, have confounded interpretation of their quality-of-care ratings.

For long-range thinking about health policy, the importance of sensitizing the medical community to costs can hardly be overstated. Physicians will face, over the next decades, mounting pressure to consider costs along with therapeutic benefits and risks (i.e., efficacy, safety) in their medical practices. Consequently, one

*The notion of providing prices in the form of actual dollar figures or relative (index) figures was considered impractical, given that the questionnaire was already lengthy and complex.

important direction in which the quality-of-care field must move is toward incorporating cost-related factors into the criteria by which care will be evaluated.

CHAPTER IV
QUALITY OF CARE IN EPISODES
OF RESPIRATORY INFECTION

As reported in Chapter II, the validation studies pointed the way to changes in the rules for creating episodes of these respiratory infections that would take greater account of the natural history of these diseases and do a better job of appropriately including or excluding services from relatively complex episodes. Chapter III documented the stages of developing the composite quality-of-care Profiles that would simultaneously take into account the presence and absence of both satisfactory and unsatisfactory elements of care. At this point, it was possible to move to the main analyses, which are the focus of this chapter.

The remainder of this chapter gives a brief description of the methods for creating episodes of respiratory conditions, for classifying levels of quality of care, for investigating differences in quality by physician type, and for analyzing whether and how quality of care might have improved between Periods I and II. Results from the various analyses are then presented, first for the five respiratory conditions other than otitis media, and second for otitis media. The latter is discussed separately because otitis media episodes were created in a manner that allowed them to coexist with episodes of the other diagnoses and because the otitis media quality-of-care criteria were more stringent than those for the other diseases.

Table IV.1
DIAGNOSES AND CORRESPONDING HICDA-I CODES
USED IN PRESENT STUDY

Name of Diagnostic Category ^a		HICDA-I Code
Streptococcal sore throat	SD ^b	34.0
Otitis media	SD	381.0, 381.2, 381.9
Sinusitis	CR ^b	461.0-.9 503.0-.9
Pneumonia (bacterial)	CR	481.0-486.9
Pharyngitis/tonsillitis	SD	462.0-.9 463.0-.9
Bronchitis	SD	489.0-490-.9
Laryngitis/tracheitis	CR	464.0-.9
Hypertrophy of tonsils	CR	500.0-.9
Influenza	SD	470.0-.9
Acute upper respiratory infection (Acute URI)	SD	460.0-.9 465.0-.9
Cough	CR	783.3

^aThese are listed in order from "most bacterial" to "most viral." Except for otitis media and bronchitis codes, they are exactly the same as those used in the earlier studies. HICDA-I codes 381.1 and 381.3-381.8 were eliminated to reduce the likelihood of confusing serous (chronic) otitis media with acute infectious otitis media. Similarly, the 491.0-.9 codes for bronchitis were eliminated to reduce any confusion between chronic bronchitis and infectious acute bronchitis.

^bSD means study diagnosis; CR means closely related diagnosis.

METHODSCREATING EPISODES OF CARE

The main steps in creating episodes of care were as follows: (1) Extract the relevant data on this cohort of Medicaid recipients in the Aid to Families with Dependent Children (AFDC) aid category onto a working tape; (2) recode claims with two diagnoses to the higher ranking ("more bacterial") diagnosis; (3) write the rules for creating episodes into an efficient computer program; (4) run a test on the first 1000 records to ensure that the programming was correct; and (5) create episodes that could then be scored according to the quality-of-care criteria. Appendix C presents the detailed steps of the methods for creating episodes. Table IV.1 presents the study and closely related diagnoses and the relevant HICDA-I codes.*

The rules for creating episodes were explicated for the programmer at some length. This section gives a very abbreviated description of the main elements of those rules. The hypothetical streams of data in Table IV.2 illustrate the major points.

*The earlier analyses (see Chapter I) indicated that (with the possible exception of DO outliers to a small degree) the several types of providers did not exhibit any shift between the two periods in the way they labelled conditions. For example, they did not seem to record a greater proportion of "more bacterial" diagnoses in Period II than they had in Period I. The EMCRO had carefully monitored diagnoses on claim forms for legibility and logic, and much of their medical review process (especially of injectable drugs) relied on diagnosis. Thus, we considered the diagnoses (which were derived directly from the claims) to be sufficiently reliable for the episode analyses.

Respiratory Infections Other Than Otitis Media

The first principle that governed creating episodes was that episodes of all study diagnoses except otitis media were created by the same basic set of rules on a single pass through the workfile. Episodes of otitis media, by contrast, were developed on a second pass through the tape and were created by a somewhat more complicated set of rules. The effect of this was to allow otitis media episodes to coexist with episodes of other respiratory infections, to allow possibly overlapping episodes to be judged by diagnosis-specific (and thus potentially dissimilar) Profiles, to allow otitis media to last six weeks (because of the initial requirement for Minimal care that a followup visit be present), and to avoid excessively complicating the programming by which other infectious illness episodes would be created. The rules for all diagnoses except otitis media are taken up first in this section.

The "general rule" was as follows. Episodes were created and recorded in the order in which they occurred for each patient.* The first visit with a study diagnosis for each patient began an episode. Generally, all visits, laboratory tests, procedures, and medications in both oral and injectable (intramuscular, or IM) form that occurred on Day 1 through Day 15 (i.e., in a two-week period) and that were given for the diagnosis of the episode or for a closely related diagnosis were assigned to the episode.** Injections, prescription drugs, and certain

*A standard rule was established to eliminate "episodes" that may have begun before the study period (see Appendix C).

**See Table IV.1. In this context, closely related would also include services for any of the other study diagnoses.

laboratory tests that occurred up to two days before the visit that initiated the episode and that met certain diagnostic criteria were also included in the episode. A special set of definitions was used so that return visits in episodes of bronchitis for children ages 0 to 7 could be recorded (because followup care was specified as an element of acceptable care in that age group).

Pharmacy claims for prescription drugs carried no diagnoses. The "prescription rule" (a slightly modified version of the original--see Chapter II) was used to assign diagnoses to pharmacy claims according to the diagnoses for visits in the episode near the date of the pharmacy claim.

Laboratory tests and other procedures typically were assigned to an episode according to the diagnosis specified on the claim. Strep and throat cultures were included in all episodes of strep throat or pharyngitis regardless of the claim diagnosis. If a laboratory test claim had no diagnosis, it was assigned by the same set of rules that governed the assignment of prescription drugs.

The simplest episode, which will account for well over half of all episodes, is illustrated by Example I in Table IV.2; in it, all services occur on a single day. Episodes involving several visits and/or prescriptions are more complex. Example II in Table IV.2 illustrates the application of some of the foregoing rules in an episode that lasts only two weeks. (It corresponds to the example of the original rules given in Chapter I.)

One complicating factor in this episode methodology (discussed in Chapter II) was the presence of "overlapping" episodes. This was

Table IV.2

HYPOTHETICAL DATA STREAMS TO ILLUSTRATE
INCREASINGLY COMPLEX EPISODES

		I. Simple Episode*					
Services →		(no other services)					
Day	→	1*					
II. Complex Episode [†]							
Services →	LT _I	V _I , I _I	V _I , P	(V ₀ P') ^{**}	15	(P'')	
Day	→ -2	1	8	(10 11)	15	(16)	
III. Complex Episode with Progression to More Bacterial Infection (I ⁺) ^{††}							
Services →	LT _I	V _I , I _I	V _I ⁺ , LT _I ⁺ , P	(V ₀) (P) (P'')	20	V _I ⁺ , LT _I ⁺	(V _I) ²⁵
Day	→ -2	1	8	(10) (15) (16)	20		

Services Symbols

V = Visit

I = Injection

LT = Laboratory test or other procedure

P, P', P'' = Prescriptions

Subscripts

I = a study or (depending on circumstances) closely related diagnosis.

I⁺ = within an episode of diagnosis I, I⁺ refers to a more bacterial study diagnosis (but not a closely related diagnosis).

0 = any other (completely unrelated) diagnosis.

* The day on which a visit occurs that begins an episode is arbitrarily designated Day 1.

** Services show in () are not assigned to episode.

[†] Based on Table 5.1, Lohr et al., 1980.

^{††} In this episode, the diagnosis progresses from a "less bacterial" infection (the visit and injection for I on Day 1) to a "more bacterial" infection (the visit, lab, and prescription for I⁺ on Day 8). Thus, the episode is extended for two weeks from Day 8, and the visit and lab test for I⁺ on Day 20 are included in the episode. The prescription on Day 15 was not recorded because the exact same drug has already been included in the episode; the prescription on Day 16 would not be included because it follows the V₀.

described as the occurrence of two or more visits in a two-to-four week period that appeared to reflect a progression of diagnoses from more to less serious conditions or, more importantly, from less to more serious conditions. ("Serious" in this context was viewed as "more bacterial.") Examples include acute URI progressing to bronchitis or pharyngitis to strep throat. This phenomenon was handled by a set of subsidiary rules that allowed for continuation of episodes in which the progression in diagnosis was from the less to the more serious condition and, for quality-of-care analyses, changed the diagnosis to the more bacterial. Example III in Table IV.2 illustrates a complex episode in which the diagnosis progresses to a more bacterial one. The major effect of the rule in this instance would be to extend the episode to (at least) Day 20 and to change the diagnosis (for quality-of-care purposes) to I+.*

Another problem (explored in Chapter II) involved services that occurred very shortly after the "end" of the two-week period for these episodes. This phenomenon was considered troublesome because, from claims data, it is often impossible to know whether these visits were elements of a very long and complex condition (which would seem to be unlikely for common, self-limited respiratory infections), whether they represented a process that could clinically and analytically be considered a second or third episode, or whether they simply were

*In the original studies, otitis media had been ranked second only to strep throat in degree of "bacterial" severity. Otitis media and strep throat are sometimes caused by different bacteria, however, and a distinction as to which might be "more" bacterial cannot be pushed very far. As noted earlier, otitis media episodes were created in a separate run and were analyzed independently. Episodes of the other study diagnoses were not extended or changed to otitis media.

miscoded by date. A rule was devised, therefore, to assign services that came shortly after the end of a two-week episode and that were for precisely for the same diagnosis as the episode (see Appendix C). Services that came shortly after the end of one episode but were for a different study diagnosis were included in the later (different diagnosis) episode. This rule was applied only in episodes that were not extended (i.e., only in episodes that did not progress from a less to a more infectious diagnosis).

Otitis Media

The same general principles were applied to otitis media. The major differences were that these episodes could last six weeks (rather than two) and that a specific set of definitions was employed by which a followup visit from either the physician giving the initial care or a different provider might be recorded for the episode. Details of the rules for otitis media can be found in Appendix C.

One HICDA-I code (381.9), for otitis media not otherwise specified, conceivably could be used for either acute infectious otitis media or chronic serous otitis media; in claims data, a definitive decision about the correct diagnosis cannot be made. This is a good illustration of the difficulties encountered when one must rely exclusively on diagnoses coded on insurance claims forms. For the present studies, claims that seemed clearly to refer to chronic otitis media were eliminated, and claims in which the billing physician resorted to the catch-all code were retained.

SCORING BY PROFILES OF CARE

The diagnosis-specific Profiles developed from the quality-of-care ratings provided by the physician panels were described in Chapter III. Four levels were defined for each age/diagnosis-specific Profile, namely, High, Acceptable, Minimal, and Unacceptable. To facilitate judging the episodes, these levels were translated into "scores" (from 1 for High to 4 for Unacceptable). (No interval level of ranking was intended by these scores). To evaluate the quality of care of the episodes, the Profiles were translated into a computer program that was merged with the tape containing the completed episodes, and a "score" for each episode was recorded.

A second set of analyses focused on reasons why certain episodes received poor quality ratings. One analysis involved the "sore throat" diagnoses (strep throat, pharyngitis) and the use of strep or throat cultures and appropriate antibiotics. A two-way table was constructed in which the cells represented the use or nonuse of such a culture and the use of appropriate or inappropriate antibiotics. Differences among provider types and between Periods I and II were investigated. Another involved otitis media, for which initially both acceptable antibiotics and a followup visit had to be present for a rating of at least Minimal. Additional frequency distributions were done that eliminated the requirements for followup visits and/or appropriate antibiotics for a Minimal score and the differences between providers and Period I and II were examined.

Table IV.3

NUMBERS OF PROVIDERS AND OF RESPIRATORY INFECTION EPISODES,
BY PROVIDER CLASSIFICATIONS

Type of Provider, Board of Certification, and Outlier Status	Period I		Period II	
	Number of: Providers ^a	Episodes	Number of: Providers ^a	Episodes
Certified Group Practice	15	561	13	329
Noncertified Group Practice	6	250	13	427
Certified MDs	60	783	57	552
Noncertified MDs				
Nonoutlier	82	1365	86	1109
Outlier	14	889	13	686
Certified DOs	5	259	9	192
Noncertified DOs				
Nonoutlier	61	2242	55	1669
Outlier	7	869	7	512
Total	250	7218	253	5476

^a Includes only providers to whom episodes were attributed.

QUALITY-OF-CARE ANALYSES

Once episodes were assigned quality-of-care scores, differences in levels of quality of care between physician types and between Periods I and II were examined. Frequency distributions were done on the number and percent of episodes in each quality-of-care level. These were run for each diagnosis (all ages together) and each provider type, in Periods I and Period II separately. For example, in Period I for all group practices, the number and percent of strep throat episodes scored as High, Acceptable, Minimal, and Unacceptable were tabulated and printed. All these data can be found in Appendices D and E.

RESULTSDISTRIBUTION OF EPISODES

The provider variables were type of practice (solo or group); type of training (Doctor of Medicine (MD) or Doctor of Osteopathy (DO)), board certification status, and "outlier" status.

Solo physicians and group practices were eligible to be in this study by virtue of having been among the more active Medicaid providers during the entire EMCRO period. Those to whom at least one episode was attributed are given in Table IV.3 for Periods I and II, along with the total number of episodes seen.* Outliers were by definition a small

*As with the earlier studies, the data were restricted to 358 providers who billed at least 100 ambulatory visits to the Medicaid program in the 1971-1973 period. Although they accounted for only about one-quarter of all Medicaid providers, they delivered 75 to 80 percent of ambulatory services in that period (see Lohr et al., 1980). The number of providers shown in Table IV.3 are lower than 358 because some providers did not give care to this AFDC cohort or for these conditions.

number of providers, although they were responsible for large numbers of episodes. Certified DOs were also few in number and they saw only a small number of episodes. The few noncertified group practices saw a small number of episodes of all conditions except pharyngitis and acute URI. The certified category included only physicians with full certification by an American specialty board; board-eligible physicians were included with the noncertified. A group practice was deemed certified if it was known or believed that, as of 1973, all or the majority of its members held full certification.* All certified group practices were composed of MDs. Outliers were as defined in Chapter I.

Subtotals of diagnosis-specific episodes can be found in Appendix D. Pharyngitis/tonsillitis and acute respiratory infection (URI) had the largest number of episodes in both periods, followed by bronchitis, influenza, strep throat, and otitis media.**

Group practices accounted for fewer episodes than did physicians in solo practice. Among the solo practice physicians, noncertified

*This determination was based on the consensus judgments of officials of the New Mexico EMCRO in 1975.

**As described in Chapter II, the present rules for creating episodes were more complex than the earlier rules and allowed for episodes to continue for longer periods. Nonetheless, the numbers of diagnosis-specific episodes in both Periods I and II were nearly the same as in the previous study for all conditions except otitis media. For otitis media, the rules underwent somewhat greater change (to give all possible opportunity for recording a followup visits). This may account in part for the lower number of otitis media episodes in the present study. The drop in the number of otitis media episodes per se (between Periods I and II) may be partly explained by the truncation of the age range for otitis media at age 8 (for reasons having to do with defining quality-of-care criteria). Because the analyses are confined to the AFDC cohort, some children may move out of the age range for otitis media, but no new children enter the cohort, thus making the number of children who might have been included somewhat lower in Period II than in Period I.

physicians outnumbered certified and accounted for the bulk of episodes.

Only a few DO providers were certified, and they saw comparatively few episodes of these conditions; thus, results involving comparisons with certified DOs are not emphasized.

A NOTE REGARDING LEVELS OF QUALITY OF CARE

As described in Chapter III, initially three levels of adequate care were defined for the diagnosis-specific profiles. These had been designed so that the requirements for High quality care were relaxed somewhat to define Acceptable care, and then those requirements were relaxed further to define Minimal care largely as a function only of use of antibiotic drugs.

The analyses reported below suggested that this approach was overly detailed. In particular, the categories of High and Acceptable appeared to cover the spectrum of appropriate care adequately, and very few episodes were scored as Minimal. Taking only the five respiratory infections other than otitis media, typically only between 2 and 6 percent of all episodes seen by groups, MDs, or DOs received scores of Minimal (using as the denominator the sum of episodes scored Minimal, Acceptable, or High).

For the remainder of the discussion about these five respiratory conditions, therefore, the Acceptable and Minimal episodes are combined into a category referred to as acceptable. The rubric "adequate care" will be taken to mean the combination of High and Acceptable/Minimal--i.e., "adequate" is the equivalent of "at least Minimal." Capitals are not used further in this chapter to distinguish

Profiles from categories in the criteria questionnaire, and all reference to high, adequate, or unacceptable care will be understood henceforth to refer to the Profiles.

DIFFERENCES IN QUALITY OF CARE AMONG TYPES OF PROVIDERS

IN PERIODS I AND II: FIVE RESPIRATORY INFECTIONS

Drawing on the previous studies, several hypotheses were articulated: (1) Group practices would deliver better care than solo providers; (2) certified practitioners would deliver better care than noncertified; (3) MDs would give better care than DOs; (4) outliers (among the noncertified) would deliver poorer care than other noncertified providers; and (5) care would improve over time. In testing these hypotheses, a simple one-tailed t-test of the differences between independent samples is used to calculate statistical significance.*

These questions are investigated by comparing, first, the percentage of episodes that received adequate quality of care and, second, the percentages that received high quality of care among the various categories of providers; Periods I and II are considered separately. Third, differences within provider types are compared between Periods I and II (in a later section).

The tables presented with the text below summarize the quality-of-care findings. Table IV.4 gives the total number and percent of the five respiratory diseases that were scored adequate (high,

*In reporting significance, any test with size of $p>0.05$ is considered not significant.

acceptable, or minimal) and that were scored high for groups and solo MDs and DOs; both Periods I and II are given. Tables IV.5 and IV.6 give the same data for more disaggregated provider categories. In these tables, "nonoutlier" and "outlier" are both categories of noncertified physicians and their sum will equal the numbers shown for all "noncertified"; the sum of certified and noncertified will equal the numbers shown for each of the three main categories of providers. Table IV.7 shows the diagnosis-specific data aggregated across all providers.

In much of this section, five respiratory infections--excluding otitis media--are considered together in the text. This is done for three reasons. First, patterns of differences among provider types do not differ in any meaningful way among the five diagnoses, so the data are aggregated for ease of presentation. Second, the otitis media episodes were created separately from episodes of the other bacterial or viral infections; they could coexist with, e.g., pharyngitis or strep throat and would neither play a role in any extensions of the length of episodes nor contribute to changes in the diagnoses of episodes. Third, the quality-of-care criteria were initially more stringent for otitis media than for the other conditions and the analyses were somewhat more detailed. Otitis media is thus discussed separately.

Group and Solo Practice

Period I

Table IV.4 shows the number and percentage of episodes with adequate (i.e., high or acceptable/minimal) and with just high quality

Table IV.4

NUMBER AND PERCENT OF EPISODES OF FIVE RESPIRATORY INFECTIONS
 WITH ADEQUATE AND HIGH QUALITY CARE, BY TYPE OF
 PROVIDER: PERIODS I AND II

Provider Type	Total Number of Episodes	Percentage of Episodes with	
		Adequate Care ^a	High Care
<u>Period I</u>			
Groups	715	67	41
MDs	2862	45	18
DOs	3230	20	7
<u>Period II</u>			
Groups	715	64	31
MDs	2279	59	28
DOs	2297	48	15

^aAdequate is the sum of high, acceptable, and minimal.

scores for the three major provider types.* In Period I, group practices had markedly higher percentages of episodes with adequate care (67 percent) than did either MDs (45 percent) or DOs (20 percent). (The one exception was bronchitis, which was treated adequately in roughly the same portion of MD and group practice episodes.) Differences between group practices on the one hand and MDs or DOs on the other were statistically significant, as was the difference between MDs and DOs ($p<0.0001$). The group practices also outstripped the solo practice physicians in the proportion with high quality care (41, 18, and 7 percent respectively). All differences were significant at $p<0.0001$.

Period II

In Period II (see Table IV.4), the general patterns, although attenuated, were the same. (The one exception was that solo MDs treated a greater percentage of their strep throat episodes adequately than did group practices.) Overall, the groups treated 64 percent of these conditions adequately, MDs 59 percent, and DOs 48 percent. The difference between group practices and MDs was significant at $p<0.01$, and the others at $p<0.0001$. Using high quality care as the metric, the groups ranked better (31 percent) than either MDs (28 percent) or DOs (15 percent) for all conditions. (For acute URI, nearly half of episodes treated by both groups and MDs received high scores.) The difference between groups and MDs was not significant; the other differences were significant at $p<0.0001$.

*Disease-specific data can be found in Appendix D: Tables D.3, D.8, and D.13 for Period I and D.16, D.21, and D.26 for Period II.

Table IV.5

NUMBER AND PERCENT OF EPISODES OF FIVE RESPIRATORY DISEASES
 SCORED ADEQUATE OR HIGH, BY BOARD CERTIFICATION AND
 TYPE OF PROVIDER: PERIODS I AND II

Board Certification and Provider Type	Total Number of Episodes	Percentage of Episodes with	
		Adequate ^a	High Care
Period I			
Certified			
Groups	484	70	44
MDs	715	55	23
DOs	230	38	10
Noncertified			
Groups	231	62	33
MDs	2147	42	16
DOs	3000	19	7
Period II			
Certified			
Groups	307	74	42
MDs	517	68	30
DOs	188	42	22
Noncertified			
Groups	408	56	22
MDs	1762	57	27
DOs	2109	49	15

^a Adequate is the sum of high, acceptable, and minimal.

Board-Certification Status

Because certified groups saw more episodes than noncertified groups in Period I and noncertified MDs and DOs more than certified solo practitioners in both periods, the possibility arises that the differences just discussed reflect being fully board-certified in some specialty, not mode of practice. Thus, one wants to know if certified practitioners in group and/or solo practice delivered better care than their noncertified counterparts. The findings were quite consistent with that hypothesis.*

Period I

In Period I (Table IV.5), certified group practices delivered better care overall than did any other category of provider. (Only for bronchitis did other providers render notably better care.) Certified MDs and certified DOs in solo practice delivered better care than their noncertified counterparts for all five conditions.

Taking the five respiratory infections together, certified group practices gave 70 percent of their episodes adequate care, noncertified groups, 62 percent ($p<0.05$). For MDs, the figures were 55 and 42 percent, respectively ($p<0.001$); for DOs, 38 and 19 percent ($p<0.001$). Both certified and noncertified groups were statistically significantly better than the respective categories of MDs and DOs ($p<0.001$) in the

*Data on specific diseases can be found in Appendix D: Tables D.1 and D.2, D.4 and D.7, D.9 and D.12 for Period I; D.14 and D.15, D.17 and D.20, and D.22 and D.25 for Period II.

percentage of episodes with adequate care, and MDs scored better than their DO counterparts ($p<0.001$).

The same patterns were observed when high quality of care was the standard. Across the five conditions,* certified groups had a higher fraction of episodes with high quality (44 percent) than did the noncertified groups (33 percent), and certified MDs a higher fraction (23 percent) than noncertified (both differences, $p<0.01$). The difference between certified and noncertified DOs (10 and 7 percent) favored the former but was not significant.

The certified and noncertified group practices had a significantly larger proportion of episodes with high quality scores than did their certified or noncertified solo practice counterparts ($p<0.001$); the difference between certified MDs and DOs and between noncertified MDs and DOs also significantly favored the MDs.

Period II

In the second period (Table IV.5), the more favorable performance by certified providers was again observed for groups and MDs, but not for DOs. Of the episodes seen by certified group practices, 74 percent received adequate care, as did 56 percent of those of the noncertified groups ($p<0.001$). The figures for certified and noncertified MDs were 68 and 67 percent ($p<0.001$). Conversely, the noncertified DOs appeared

*For some diseases (see Appendix D), noncertified practitioners tended to give the same or better care than certified, but except for pharyngitis among solo physicians these differences were most likely artifacts of small numbers.

to give somewhat better care than the certified DOs (49 percent with adequate care compared to 42 percent), but the difference was not significant.

The certified groups gave adequate care more often than their certified solo practice peers, but significantly so only with respect to DOs ($p<0.001$). Noncertified groups gave slightly worse care than noncertified MDs, and somewhat better care than noncertified DOs ($p<0.05$). Both certified and noncertified MDs gave better care than their DO peers ($p<0.001$).

When the metric is the percentage of episodes receiving high quality of care, the pattern of certified physicians giving better care than noncertified physicians was again observed, but significantly so only for groups and DOs ($p<0.01$). (The exceptions involved bronchitis and acute URI for MDs and strep throat for DOs.)

Certified groups had a higher fraction of episodes of high quality care than did certified MDs or DOs ($p<0.001$). Noncertified groups gave high quality care to a slightly lower proportion of episodes than did MDs ($p<0.05$) and to a slightly higher proportion than did DOs ($p<0.01$). MDs in both categories were better than DOs (certified, $p<0.05$; noncertified, $p<0.001$).

Summary

If one were to evaluate the factors predictive of better care, group practice and board certification status both are important at the "baseline" (i.e., in Period I), and the differences are greater between the group and solo practice physicians than they are between the

certified and noncertified physicians. In Period II, the influence of being in a group practice has diminished but not disappeared; board certification, however, remains an important predictor of better quality of care. The differences between certified and noncertified practitioners is greater than the differences between groups and solo practitioners, and in fact the presumed "benefit" of being in a group practice is not observed among noncertified providers. From "best" to "worst," the ranking (by percentage of episodes receiving high care) is certified groups, certified MDs, noncertified MDs, noncertified groups, and certified DOs, and noncertified DOs.

The question remains as to whether type of physician training (namely, medical or osteopathic) had an independent effect. Data aggregated across the five infections suggest that it does. As already noted, comparing like categories of MDs with DOs showed that MDs consistently gave better care. Moreover, in both Periods I and II, the noncertified MDs gave better care than the certified DOs (using both adequate and high quality care as the standard). Although the certified DOs saw many fewer episodes than did the noncertified MDs, the difference in the proportion receiving adequate care was statistically significantly in favor of the noncertified MDs, and higher but not significantly so for the proportion receiving high quality care.

In summary, the following conclusions were drawn: (1) Board certification was a major predictor of better quality care for any type of provider in both periods; (2) in both periods, within the board-certified category, group practices were more likely to render superior care than were solo practice physicians; (3) in the later

period, this positive effect of being in a group practice (compared with being in solo practice) was not observed for the noncertified physicians; and (4) among the solo practitioners, being an MD was an important predictor of better care.

Outlier Status

Outliers, it will be recalled, were the very few physicians who in the first study had been identified as accounting for a large fraction of injectable drugs denied by the EMCRO for medical reasons. All were noncertified, and all were in solo practice. The earlier studies had shown that outliers consistently gave poorer care than other types of physicians, at least as measured by the proportion of episodes receiving inappropriate antibiotics in oral and injectable forms. The hypothesis is that, even with more comprehensive criteria as the metric, they will still give poorer quality care. In the discussion below, "nonoutliers" refers only to noncertified physicians in solo practice.

Period I

As seen in Table IV.6, nonoutlier physicians among both the MDs and DOs gave substantially better care than outliers.* Among MDs, nonoutliers gave 56 percent of their episodes adequate care; outliers, 21 percent ($p<0.001$). Respective figures for the DOs were 24 and 7 percent ($p<0.01$). The same general patterns were observed for the

*See Appendix D for disease-specific data: Tables D.5 and D.6, D.10 and D.11 for Period I; Tables D.18 and 19, D.23 and D.24 for Period II.

Table IV.6

NUMBER AND PERCENT OF EPISODES OF FIVE RESPIRATORY DISEASES
 SCORED ADEQUATE OR HIGH, BY OUTLIER STATUS AND
 TYPE OF SOLO PROVIDER: PERIODS I AND II

Provider Type ^a and Outlier Status	Total Number of Episodes	Percentage of Episodes with		
		Adequate ^b	High Care	
Period I				
MDs				
Nonoutlier	1294	56	22	
Outlier	853	21	8	
DOs				
Nonoutlier	2150	24	9	
Outlier	850	7	3	
Period II				
MDs				
Nonoutlier	1088	62	29	
Outlier	674	48	24	
DOs				
Nonoutlier	1610	55	16	
Outlier	499	31	9	

^aAll are noncertified solo practice physicians.

^bAdequate is the sum of high, acceptable, and minimal.

proportions of episodes receiving high quality care, and the differences were again significant ($p<0.001$).

Both MD classifications delivered adequate care to higher proportions of episodes in Period I than their DO peers ($p<0.001$). This was true for all conditions at an adequate level of care and for all conditions except strep throat and pharyngitis at a high level of care.

Period II

The pattern of nonoutliers' typically giving better care than outliers carried into Period II (Table IV.6). (The only exception was acute URI among MDs.) Among MDs, the percentages of episodes receiving adequate care were 62 percent for nonoutliers and 48 percent for outliers ($p<0.001$); among DOs, the respective figures were 55 and 31 percent ($p<0.001$). MDs in both classifications gave better care than DOs ($p<0.001$).

The same trends were observed for the proportions of episodes receiving high quality care in Period II (Table IV.6). (The exceptions were bronchitis and acute URI (MDs) and strep throat and pharyngitis [for small numbers of cases among DOs].) For nonoutliers and outliers, respectively, the percentages of all episodes that received high quality care were 29 and 24 percent for MDs and 16 and 9 percent for DOs (both $p<0.05$). As expected, the differences between nonoutlier MDs and DOs and outlier MDs and DOs in the proportion of episodes receiving high quality care all significantly favored the MDs ($p<0.001$).

Taking all conditions together, in both Periods I and II the nonoutlier DOs typically gave somewhat better care than the MD outliers. In Period II, however, the MD outliers gave a higher proportion of episodes high quality care than did the nonoutlier DOs (because of their better care for bronchitis and acute URI by the MDs).

In the previous section, it was noted that certification played an important role in predicting better care. Interestingly, however, the findings for nonoutlier noncertified physicians, especially MDs, were not meaningfully different from the results for certified physicians. Some persons might argue, therefore, that certification is really not important. Persons seeking care, however, cannot know in advance whether any specific physician from the pool of noncertified physicians is likely to be a "nonoutlier." For this reason, these results are still interpreted as suggesting that seeking care from physicians who are board-certified is likely to be an important factor in acquiring higher quality of care than one might otherwise obtain.

DIFFERENCES IN QUALITY OF CARE BETWEEN PERIODS I AND II

In all previous studies, care for these specific conditions and care generally (as judged by the use of appropriate injectable drugs) improved between Periods I and II. This was true mostly for the use of inappropriate antibiotics, which was the area in which the EMCRO made its greatest efforts. When use of prescription antibiotics (which was not reviewed by the EMCRO) was included in the criteria, the improvement over time was diluted somewhat (Lohr et al., 1980).

Using those findings as guides, one might predict that (1) care

would improve between Periods I and II, but that (2) because several other elements have now been incorporated into the quality-of-care criteria, the improvement will be somewhat attenuated (compared with the earlier studies). The hypothesis to be tested is that, for all types of providers, the proportion of episodes receiving adequate care will be higher in Period II than in Period I. Because of the way in which the quality-of-care profiles are constructed, improvement in the use of injectable drugs (particularly antibiotics) would be expected to affect the proportions receiving only minimal or acceptable care, but not necessarily high quality care; thus, the expected effect is that the proportion of episodes receiving high quality care, although higher in Period II than Period I, will not have risen as much as the proportion receiving only minimal or acceptable care.

By Diagnosis

The percentages of episodes receiving high, acceptable/minimal, and unacceptable care were calculated for both Periods I and II. A total of 6807 episodes of the five diagnoses were treated in Period I; of these, 1041 (15 percent) received high quality care, 1400 (21 percent) received acceptable/minimal care; and 4366 (64 percent) received unacceptable care. By contrast, of the 5291 episodes seen in Period II, 1198 (23 percent) received high quality care, 1723 (33 percent) received acceptable/minimal care, and 2370 (45 percent) received unacceptable care. The increases for both high and acceptable/minimal were highly significant ($p<0.0001$).

Table IV.7

NUMBER AND PERCENT OF EPISODES OF FIVE RESPIRATORY INFECTIONS,
BY LEVEL OF CARE: PERIODS I AND II

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	No.	%	No.	%	No.	%	No.	%	No.	%
<u>PERIOD I</u>										
High	44	11	61	3	225	15	124	24	587	28
Acceptable/Minimal	74	19	874	38	373	25	14	3	65	3
Adequate	118	30	935	40	598	40	138	27	652	32
Unacceptable	273	70	1380	60	917	60	377	73	1419	68
Total	391	100	2315	100	1515	100	515	100	2071	100
<u>PERIOD II</u>										
High	53	11	130	6	348	35	108	35	559	41
Acceptable/Minimal	161	34	1072	49	396	40	26	8	68	5
Adequate	214	45	1202	55	744	75	134	43	627	46
Unacceptable	259	55	970	45	244	25	175	57	722	54
Total	473	100	2172	100	988	100	309	100	1349	100

Care improved for all five conditions (see Table IV.7). By Period II, the percentages of episodes receiving adequate care (aggregated across all physicians) were as follows: strep throat, 45 percent; pharyngitis, 55 percent; bronchitis, 75 percent; influenza, 43 percent; and acute URI, 46 percent. Care for bronchitis improved the most, largely as a function of the decrease in use of contraindicated drugs. Although improvements of this magnitude are certainly meaningful, even with them the average level of care for most of these conditions remained deplorably low.

The percentages of episodes receiving high quality care also either rose or (in one condition) stayed the same, and by Period II the figures were as follows: strep throat, 11 percent; pharyngitis, 6 percent; bronchitis, 35 percent; influenza, 35 percent; and acute URI, 41 percent. Interestingly, for influenza and acute URI in both Periods I and II, if care was appropriate it was considerably more likely to be high than acceptable or minimal. Almost exactly the opposite was true for strep throat and pharyngitis, in which appropriate care was more likely to be acceptable or minimal than high (especially for pharyngitis).

By Type of Providers

Between Period I and II, the percentage of episodes receiving adequate care decreased slightly for groups and increased for MDs and especially DOs (both $p < 0.0001$). The proportion of the groups' episodes receiving high quality care decreased significantly ($p < 0.01$), whereas the proportion of MDs' and DOs' episodes receiving that level of care rose ($p < 0.001$).

Certified group practices improved between Periods I and II in the percentage of episodes receiving adequate care; hence, the decline for all groups was attributable to poorer care from noncertified group practices in Period II, especially for pharyngitis. (The number of pharyngitis episodes seen by noncertified groups rose in Period II without a concomitant increase in the use of throat cultures; use of such cultures had been a criterion for acceptable and high quality care in pharyngitis.) All other comparisons for subgroups of providers (i.e., for certified and noncertified MDs and DOs, and for outliers and nonoutliers) showed that care improved significantly ($p<0.01$) between Periods I and II in the proportion of episodes receiving adequate care and high quality care.

COMPARISONS BETWEEN THE EARLIER AND PRESENT STUDIES

The proportions of episodes receiving satisfactory or appropriate care in Periods I and II were fairly similar between the earlier study (Lohr et al., 1980) and the present one. Appropriateness in this instance means appropriate use of injectable and oral antibiotics in the earlier study, and adequate (high, acceptable and minimal) care as judged by the comprehensive criteria in the present study. In Period I, the proportions of episodes (of all conditions) judged to have received satisfactory care were lower in the present study than in the earlier one. In Period II, compared with the results of the earlier study the proportions of episodes judged appropriate were higher for strep throat, about the same for pharyngitis, and somewhat lower for bronchitis, influenza, and acute URI.

Of all the differences between the previous and the present study in percentages of episodes receiving appropriate care, only the poorer showings (in the present study) for bronchitis (in Period II) and for acute URI (Period I) would seem to have any real importance. The explanations may arise from the greater complexity of the criteria and the rules for creating episodes. "Searching" for the followup visit in bronchitis (for children younger than 8 years) and assigning it to an existing episode may have had the effect of eliminating single-visit episodes of bronchitis that received no antibiotics (and would previously have been counted as episodes with appropriate care). Moreover, some episodes of the nonbacterial infections (e.g., acute URI) were "extended" (under certain conditions) into episodes of more bacterial infections such as pharyngitis and strep throat. This would remove such episodes (which, if they involved no antibiotics, would previously have been considered appropriate) and leave behind only nonbacterial episodes that were at greater risk of being considered inappropriate (because they may have involved the use of antibiotics).

MANAGEMENT OF SORE THROAT EPISODES

Another summary measure of differences in quality of care was the number of episodes of "sore throat" (strep throat and pharyngitis) with four combinations of possible diagnosis/treatment modalities. These combinations were as follows: (1) both a throat culture and appropriate antibiotics; (2) a throat culture but no appropriate antibiotics; (3) appropriate antibiotics but no throat culture; and (4) neither a throat culture nor appropriate antibiotics. Appropriate antibiotics refer to

Table IV.8

PERCENT OF SORE THROAT EPISODES RECEIVING MOST AND LEAST
 SATISFACTORY CARE, BY BOARD CERTIFICATION STATUS AND
 TYPE OF PROVIDER: PERIODS I AND II

Board Certification and Provider Type	Total Number of Sore Throat Episodes	Percentage with Most Satisfactory Care ^a	Percentage with Least Satisfactory Care ^b
Period I			
Certified Groups	199	26	19
MDs	265	2	44
DOs	158	0	70
Noncertified Groups	81	11	22
MDs	1039	2	61
DOs	964	>1	79
Period II			
Certified Groups	185	25	20
MDs	247	16	22
DOs	113	4	80
Noncertified Groups	252	13	31
MDs	841	6	43
DOs	1007	3	51

^aAppropriate antibiotics and a throat culture used.

^bNeither appropriate antibiotics nor a throat culture used.

Table IV.9

PERCENT OF SORE THROAT EPISODES RECEIVING MOST AND LEAST
 SATISFACTORY CARE, BY OUTLIER STATUS AND
 TYPE OF PROVIDER: PERIODS I AND II

Outlier Status and Provider Type	Total Number of Sore Throat Episodes	Percentage with Most Satisfactory Care ^a	Percentage with Least Satisfactory Care ^b
Period I			
MDs			
Nonoutlier	567	3	41
Outlier	472	1	84
DOs			
Nonoutlier	717	>1	76
Outlier	247	0	87
Period II			
MDs			
Nonoutlier	491	10	30
Outlier	350	>1	61
DOs			
Nonoutlier	707	4	42
Outlier	300	0	71

^aAppropriate antibiotics and a throat culture used.

^bNeither appropriate antibiotics nor a throat culture used.

the therapies allowed for a score of at least minimal care for the two diseases; "no antibiotics" and "no culture" together is allowed at the minimal level for pharyngitis but not strep throat.

Tables IV.8 and IV.9 give the numbers and percentages of sore throat episodes with most and least satisfactory care--i.e., (1) and (4) above.* Certified groups gave the most satisfactory care in the highest percentage of sore throat episodes in both periods (a throat culture and appropriate antibiotics) and the least satisfactory care (neither a throat culture nor appropriate antibiotics) in the lowest percentage of episodes. In Period I, the noncertified groups appeared to rank second; in Period II, the certified MDs had the second-best record. For all other providers in Period I and for all DOs and the outlier MDs in Period II, the patterns of care were very discouraging (in the sense of not using the most satisfactory mode of treatment and using the least satisfactory mode).

By and large, the improvements between Period I and II are explained by improved use of antibiotics, not by any greater use of throat cultures. This is entirely expected, of course, because of the EMCRO's focus on injectable antibiotics and lack of any systematic effort to expand the use of diagnostic procedures. The one exception appeared to be among noncertified DOs (especially nonoutliers), whose use of throat cultures (without appropriate antibiotics) in these two conditions rose over time more than the use by any other type of provider.

*Detailed data on all four categories can be found in Appendix E.

DIFFERENCES IN QUALITY OF CARE AMONG TYPES OF PROVIDERSIN PERIODS I AND II: OTITIS MEDIA

As noted earlier, otitis media was investigated independent of the other conditions. Otitis media episodes were created separately and were not part of the process by which episodes might be "extended" as diagnoses changed to more bacterial ones. Rather, otitis media episodes were allowed to "coexist" or overlap with those of the other conditions. This allowed such episodes to be given quality-of-care scores according to diagnosis-specific criteria.

Initially, the criteria for the minimal level of care in otitis media were more stringent than for any other condition, because a followup visit was required. The earlier studies suggested that no more than about 20 percent of the episodes of otitis media would have a followup visit, and even though the present specifications for followup visits were somewhat broader than those used earlier, nonetheless a substantial fraction of the episodes were expected not to meet the minimal criteria.

One problem with acute otitis media is the ambiguity of the diagnosis itself on Medicaid claim forms. One HICDA-I code for this condition (391.9) can be and probably was used as a default, such that episodes of chronic otitis media might be included in that category. Typically, chronic (nonsuppurative) otitis media does not require antibiotic treatment (although the literature reflects some controversy on this point), and a followup visit (although desirable) might not be required to take place within a four-to-six week period after the initial visit of that diagnosis.

For these reasons, the initial minimal-level criteria for otitis media (which required appropriate antibiotics and a followup visit) were moderated into two other options.* The first modification was that appropriate antibiotics would still be required but that no followup visit would be required for minimal care. The second modification was the reverse, namely, that a followup visit would be required, but that antibiotics would not be required (although appropriate antibiotics would still be allowed).**

Period I: Adequate Care

As expected, based on the most stringent (original) criteria, the percentage of episodes receiving adequate care in Period I was extremely

*The requirements for acceptable and high quality care involved mainly the absence of injectable medications (such as analgesics or antihistamines) and of certain diagnostic procedures (such as chest x-rays). The modifications for a minimal score did not affect the requirement for an acceptable or high score.

**A third modification was considered as well: neither a followup visit nor antibiotics would be required (although appropriate antibiotics would be allowed). This very liberal modification would be expected to produce the best scores for all providers, as indeed was observed for all but DO outliers. Nonetheless, it is difficult to defend these extremely liberal criteria in the absence of more complete data on these children (e.g., from medical records). Certainly not all of these cases could possibly have been chronic otitis media; moreover, although one might be comfortable either with allowing no antibiotics to be used or with not requiring a followup visit, care of acute otitis media in which both of these elements are missing would be deficient according to all commonly accepted practices today. Thus, although these liberal criteria do give physicians the most "benefit of the doubt," they were finally dismissed as simply too lax and this option was not considered further.

low (see Table IV.10).* Better quality-of-care scores resulted when the criteria required only appropriate antibiotics (and not a followup visit); even so, only about one-quarter of all otitis media episodes in Period I received satisfactory care.

When the otitis media episodes were disaggregated by provider type (see Tables IV.11 and IV.12), the sample sizes became very small, so the findings were mainly impressionistic. With both the original and modified criteria, MDs gave better care than did group practices or DOs, and the certified physicians or groups better care than noncertified (except for groups when the modified criteria were used). Among the noncertified solo practitioners (see Table IV.12), the nonoutliers gave better care than the outliers; interestingly, the MD outliers had better records than the DO nonoutliers.

Period II: Adequate Care

The care for otitis media was generally better in Period II than in Period I. The EMCRO's efforts were directed entirely at the use of injectable drugs and not at improved followup care; as noted, no more than about a one-quarter of the episodes on average could be expected to have a followup visit. Thus, the level of improvement using the original criteria would be expected to be only moderate, as was indeed the case (Table IV.10). On the basis of the more liberal criteria (appropriate antibiotics were required but a followup visit was not),

*See Appendix D Tables D.27-D.32 for the detailed data on which these findings are based. The modification of the minimal criteria that allowed, in effect, a visit to be substituted for antibiotic use (by requiring followup but not antibiotics) did not produce better quality-of-care scores than requiring the antibiotics but not the visit. Consequently, the second modification was not considered further.

Table IV.10

NUMBER AND PERCENT OF EPISODES OF OTITIS MEDIA,
BY LEVEL OF CARE: PERIODS I AND II

Level of Care	Period I		Period II	
	No.	%	No.	%
<u>Initial Criteria^a</u>				
High	24	6	13	7
Acceptable/Minimal	15	4	17	9
Adequate	39	9	30	16
Unacceptable	374	91	155	84
Total	411	100	185	100
<u>Modified Criteria^a</u>				
High	24	6	13	7
Acceptable/Minimal	90	22	80	43
Adequate	114	28	93	50
Unacceptable	297	72	92	50
Total	411	100	185	100

^aA followup visit and appropriate antibiotics are required for minimal care.

^bA followup visit is not, but appropriate antibiotics are, required for minimal care; see text.

percent. (Both overall improvements were significant at $p<0.001$).

the percentage of episodes receiving adequate care rose sharply to 50

Differences by provider type in Period II were somewhat anomalous, and they were sensitive to which criteria set was used (see Tables IV.11 and IV.12). Based on the most stringent criteria, certified groups gave the best care, and certified MDs the next best care. When the modified criteria were used as the standard, certified providers and noncertified solo providers were judged to have provided good care in a substantial proportion of episodes.

Based on the original criteria, improvements between Periods I and II were not significant for any provider classification except certified group practices ($p<0.05$); when the modified criteria were used, they were significant ($p<0.05$) for certified groups, outlier MDs, outlier DOs, and nonoutlier DOs.*

The general impressions from these otitis media results supports the observations reported earlier: being board certified and being an MD were likely to be predictive of better care, but being in a group practice did not by itself augment the likelihood of giving good care for otitis media. One explanation is that groups may have seen relatively more chronic otitis media than other providers, including

*Had the least rigorous criteria had been used (i.e., requiring neither a followup visit nor appropriate antibiotics), 63 percent of the group practice episodes, 65 percent of the MD episodes, and 74 percent of the DO episodes would have been scored as having been given adequate care. These, too, are all improvements compared with the pertinent figures for Period I. Because appropriate antibiotics were allowed in this most lenient modification to the original profile, the only element of care that would give rise to an unacceptable score was the use of inappropriate antibiotics. This means that, in Period II, at least one-quarter to one-third of the episodes (depending on provider type) received inappropriate antibiotics. These figures are in line with the findings reported in the previous study.

Table IV.11

NUMBER AND PERCENT OF OTITIS MEDIA EPISODES WITH ADEQUATE CARE,
 CERTIFICATION STATUS AND TYPE OF PROVIDER:
 PERIODS I AND II

Board Certification and Provider Type	Total Number of Episodes	Percentage with Adequate Care by:	
		Original ^a Criteria	Modified ^b Criteria
Period I			
Certified			
Groups	77	13	22
MDs	68	18	38
DOs	29	7	41
Noncertified			
Groups	19	5	32
MDs	107	9	36
DOs	111	4	13
Period II			
Certified			
Groups	22	41	55
MDs	35	20	40
DOs	4	0	50
Noncertified			
Groups	19	16	26
MDs	33	15	64
DOs	72	8	54

^aA followup visit and appropriate antibiotics are required for minimal care.

^bA followup visit is not, but appropriate antibiotics are, required for minimal care.

Table IV.12

NUMBER AND PERCENT OF OTITIS MEDIA EPISODES WITH
ADEQUATE CARE, BY OUTLIER STATUS: PERIODS I AND II

Provider Type and Outlier Status	Total Number of Episodes	Percentage with Adequate Care by:		
		Original ^a Criteria	Modified ^b Criteria	
<u>Period I</u>				
MDs				
Nonoutlier	71	11	41	
Outlier	36	6	28	
DOs				
Nonoutlier	92	4	15	
Outlier	19	0	0	
<u>Period II</u>				
MDs				
Nonoutlier	21	24	62	
Outlier	12	0	67	
DOs				
Nonoutlier	59	7	58	
Outlier	13	15	38	

^aA followup visit and appropriate antibiotics are required for minimal care.

^bA followup visit is *not*, but appropriate antibiotics are, required for minimal care.

cases perhaps referred from solo practice physicians, but that these visits were coded with an ambiguous diagnostic code and were thus ascribed to acute otitis media. As noted previously, chronic and acute otitis media are associated with different therapeutic needs.

Despite the problems noted about interpreting the otitis media data, the findings do suggest that the general level of care delivered for this condition in this Medicaid population was not particularly laudable. Probably no more than half of these episodes received reasonably appropriate care. Again, this was the observation for the other five respiratory infections, which cover a spectrum from relatively minor respiratory ailments to bacterial infections with potentially serious long-term sequelae.

Periods I and II: High Quality

The criteria defining high quality of care for otitis media were not modified in any way; they required a followup visit, appropriate antibiotics, and the absence of certain injectable drugs and irrelevant laboratory tests (see Chapter III). In both Periods I and II (see Table IV.10), the percentages of episodes receiving a high quality-of-care score were very low.

In Period I, certified MDs had the highest percentage of episodes with high quality of care (15 percent); overall, MDs (8 percent) were higher than group practices (6 percent) or DOs (3 percent). Outliers had no episodes scored high. In Period II, certified group practices had the highest percentage with high quality of care (18 percent), followed by certified and nonoutlier noncertified MDs (14 percent each).

Overall, the groups and MDs had the same record for high quality of care (12 percent), compared with 0 percent for DOs. Generally, the proportion of episodes receiving high quality from groups and MDs rose, whereas the DOs recorded no such improvement.* No improvement was significant.

SUMMARY

The major findings can be summarized as follows. Board-certified physicians in both group and solo practice gave better care than noncertified. Except for otitis media, certified groups (all of whom were MDs) gave the best care overall, suggesting that mode of practice may have an additional beneficial effect on quality of care. MDs consistently gave better care than DOs, both among the certified and noncertified; taking the five respiratory infections together, noncertified MDs gave better care than certified DOs. Among the solo practice physicians who were not board certified, better care was recorded for nonoutliers than for outliers. Outlier MDs gave somewhat better care than nonoutlier DOs in some instances. Thus, among the solo practice physicians, being an MD had an additional beneficial effect on quality of care.

*The only significant difference ($p<0.05$) among providers was that noncertified MDs had a better record than noncertified DOs in Period II.

DISCUSSIONFACTORS LEADING TO GOOD AND BAD CARE

Whether group practices consistently give better care than physicians in solo practice is still an open question. Some literature suggests that they do; for instance, certain prepaid group practices or the old OEO neighborhood health centers serving disadvantaged populations may have delivered somewhat better care than did other sources of care available at the time (see, e.g., Brook and Williams, 1975). Certainly the claim that better technical care is available in group settings is still widely accepted as correct, especially with respect to the process of medical care. Regarding outcomes, one recently published work (Dutton and Silber, 1980) reported that children who used solo practitioners tended to have higher-than-expected rates of common ambulatory illnesses, and children who used fee-for-service or prepaid group practices or hospital outpatient departments tended to have the same or lower-than-expected rates.*

Group practices may well give better medical care and have better patient outcomes, but the elements that might explain such observations (e.g., selection effects of doctors into groups, subconscious or at least very informal "peer review" within the group) are still obscure. Results of this study also suggest that group practice may be a factor in promoting better quality of care, given that in both Periods I and II such practices altogether gave better care than any other type of

*The authors noted that some differences were small and that not all potential confounding factors were (or could be) accounted for. Hence, they describe their findings as suggestive, not conclusive.

provider. This interpretation must be tempered by the fact that, by and large, this was attributable mainly to the certified groups; as noted earlier, they were not typical fee-for-service group practices because they were mostly large, multi-specialty, and academically based. Thus, further study is needed to explicate whether group practice is an important determinant in quality of care and, if so, why.

Presumably because of the more demanding requirements for a high score for strep throat and pharyngitis (both requiring throat cultures) and for otitis media (requiring a followup visit), if care was adequate at all it was more likely to be acceptable or minimal than high. Conversely, adequate care for influenza and acute URI was more likely to be high (than acceptable or minimal), perhaps because they did not have "extra" requirements for the performance of some service in order that quality of care be considered high. In other words, a judgment that quality of care is good might be more likely when not using services is the more prominent element in the evaluation than when using specified services is the more important element. This in turn underlines the need for peer review organizations to develop criteria by which underservice, as well as overservice, can be evaluated.

A "QUASI-CONTINUUM" OF CARE

This study indicated that defining three profiles of adequate care--namely, high, acceptable, and minimal--was feasible. Such profiles might, in fact, be readily convertible into a "quasi-continuum" by which to evaluate quality of care. This would be especially useful in situations involving more complex diseases and/or more elements of care.

Interestingly, however, the three-tiered scheme proved to be unnecessarily detailed, at least for these types of analyses with these relatively simple diseases. Virtually all episodes that were scored as adequate received either high or acceptable scores, and relatively few received only minimal.

One explanation may be that care for common respiratory conditions of this nature is not all that complex or elaborate; therefore, only a few elements of care influence the score that an entire episode of care might receive. For most conditions, for example, use of certain oral symptomatic medications (e.g., non-narcotic analgesics), certain laboratory tests (e.g., urinalysis), and followup visits played no role in determining whether care was at least adequate (although they may have influenced whether care was scored high). The conclusion is that defining two levels of satisfactory care probably suffices for this type of quality-of-care study.

The interpretation that only two levels of satisfactory care are needed in such profiles may not be generalizable to more complicated acute conditions or chronic conditions. For serious acute illnesses and chronic diseases, for instance, one might wish to develop profiles that incorporated "art-of-care" measures. These in turn might figure importantly in trying to define an additional level of, e.g., superior quality. If these humanistic elements were integrated into a quality-of-care continuum, more than one category of unacceptable quality might also be defined. Additionally, for chronic diseases, the length of an episode is typically much longer than for even serious acute illnesses. One might wish to define profiles of care that would

reflect sustained versus sporadic high (or low) quality of care over that longer period of time.

Such refinements would require additional research within the quality-of-care assessment field; more detailed, valid, and reliable data; and better methods of aggregating and manipulating such data. Whether such improvements can or will be forthcoming any time soon, especially within the PSRO program, is largely a matter of conjecture. Much of the emphasis within PSROs remains on cost-containment and on inpatient care, and a deliberate shift in emphasis to quality assessment in ambulatory care is not on the immediate horizon. Nevertheless, cost (or utilization) review and quality review cannot indefinitely remain estranged, and we may hope that PSROs and other agencies now embroiled in cost-oriented review will eventually respond to the need for developing better measures of quality of care.

IMPROVED CARE VERSUS LINGERING INADEQUACIES

The percentages of episodes receiving adequate care (the sum of episodes scored minimal, acceptable, or high) and the percentages receiving high quality of care both rose between Periods I and II. The patterns were interpreted as suggesting that care went from poor to acceptable levels at least as much as (or more than) it went from one acceptable category to a higher one. For disadvantaged persons, this is surely a more meaningful outcome than if poor care (especially as widespread as it had been initially) had remained poor and relatively good care (as restricted as it was) had improved.

The physicians giving the poorest care initially (e.g., the outliers) improved the most between the two periods. With only trivial

exceptions, care given by the better providers did not decline. These results also lend credence to the assertion that the variance in quality of care declined. From the point of view of equity for this Medicaid-eligible population vis-a-vis nondisadvantaged populations, this is a worthwhile accomplishment of the EMCRO.

Despite this optimistic interpretation, for most conditions the care given in Period II was adequate only in about half the episodes. Even given the superior care delivered by some provider types (especially the certified group practices), for some conditions the general level of care to this disadvantaged population was still rather dismal. Thus, regardless of past achievements of peer review organizations or the level of resources available, the implication is that PSROs must continue to be alert and responsive to quality-of-care issues.

CHAPTER V
CONCLUSIONS AND POLICY IMPLICATIONS

This study had four principal objectives: to develop valid rules for creating episodes of care from computerized claims data, to develop comprehensive age- and diagnosis-specific profiles of quality of care, to judge the quality of care for several respiratory infections, and to investigate differences in quality of care by characteristics of physicians. The respiratory illnesses studied were streptococcal sore throat, pharyngitis and/or tonsillitis, acute otitis media, acute bronchitis, influenza, and acute upper respiratory tract infection. (URI). Some of this work extends or refines the original episode analyses reported in Lohr et al. (1980).

Data for this study were taken from the New Mexico Medicaid program on a cohort of persons who had been enrolled in the Aid to Families with Dependent Children (AFDC) aid category continuously for the years covered by the study. The cohort changed in no way except natural aging. It comprised primarily women and children, was predominantly white, and accounted for a large portion of New Mexico Medicaid expenditures during this period.

Quality-of-care profiles were based on explicit process criteria and took into account both the presence and absence of appropriate and inappropriate elements of care. Included in the profiles were use of intramuscular (IM) and oral antimicrobials (antibiotics) singly and in numerous combinations, IM and oral medications typically used for

symptomatic relief of these respiratory infections, a variety of laboratory and diagnostic tests, and followup visits.

Quality of care was evaluated on the basis of these profiles and as a function of the "structural" characteristics of the care-giver. The characteristics of interest were type of practice organization (group practice versus solo practice), type of physician training (MD [medicine] versus DO [osteopathy]), board certification status,* and "outlier" status. The latter referred to a small set of physicians, identified in the first of this series of studies, who had given poor care as judged by their inappropriate use of injectable drugs of all types.

One element of this study was to examine the effect of the New Mexico Experimental Medical Care Review Organization (EMCRO) on use of services (especially injectable drugs) and quality of care. The EMCRO had instituted peer review of ambulatory care delivered to this Medicaid population in 1971 and continued such review into 1975. The effect of the EMCRO quality-of-care efforts was assessed over the first two years of their operation, using a quasi-experimental approach involving "before" and "after" periods that corresponded, respectively, to the time before any significant educational or review activities took place and the period after the major impact of the EMCRO had occurred.

*Certified physicians were those with full certification from an American specialty board; "board-eligible" physicians were included with the noncertified. Group practices were classified as certified if all or a majority of their members were certified.

EMCROs, although not so designed, have been considered prototypical of areawide peer review organizations such as Professional Standards Review Organizations (PSROs). Many aspects of the PSRO program (typically those related to inpatient rather than ambulatory care review) were modeled on certain EMCRO organizational features. The New Mexico EMCRO also provided a good standard for ambulatory care review, based on the procedures it developed for examining claims for services billed under the Medicaid program. Thus, examination of the effects of the New Mexico EMCRO should provide a useful guide as to how and to what degree such organizations can improve the quality of ambulatory medical care. Although generalizations from the New Mexico EMCRO to the national PSRO program should be made cautiously (because of several factors that made that EMCRO's experience relatively unusual), recent work in settings such as New York City (Paris et al., 1980) tend to confirm these findings, thereby strengthening the argument that the New Mexico studies do provide reliable and policy-relevant results.

PRINCIPAL FINDINGS

DETERMINING EPISODES OF CARE

1. The original computer-based "New Mexico" rules for defining episodes of care were validated against external and quite complete claims data for ambulatory care taken from the Rand Health Insurance Study (HIS). The episodes created by the original rules were identical to "true" HIS episodes about 80 percent of the time. The major problem in the remaining 20 percent of episodes was that services occurring after the end of a two-week episode were not included in the episode.

2. Various modifications of the rules and clinical guidelines (for assigning services in long and complex sequences to individual episodes) resulted in a set of rules that improved the match between the true HIS episodes and those defined by the New Mexico rules to about 93 percent. In the remaining 7 percent of episodes, the principal error arose from not correctly including or excluding oral medications (typically antibiotics) for which diagnoses were ambiguous or missing. Because this happened in less than 5 percent of all episodes, it was considered a problem well within a tolerable level of "noise" for a methodology of this sort.

3. Almost one-quarter of the HIS episodes used to validate the New Mexico rules had double diagnoses. These were either "progressions," defined as episodes in which the respiratory infection appeared to progress from a less bacterial to a more bacterial diagnosis (e.g., from acute URI to pharyngitis or tonsillitis), or "combinations," defined as two respiratory infection diagnoses appearing on the claim for the physician visit beginning an episode. To deal with the first problem, the rules were modified to include provisions for extending the time interval of episodes in which a progression occurred and changing the diagnosis of the episode to the more bacterial one. To deal with the second, the "hierarchical" approach (always choosing the more bacterial of the two diagnoses) was retained. This had been in effect for the earlier studies, and no better approach (conceptually or methodologically) could be devised for dealing with claims of this sort.

4. The total number of episodes defined from the New Mexico Medicaid data was 7218 in Period I and 5476 in Period II. They were distributed by diagnosis as follows (Periods I and II, respectively)--

strep throat: 391, 473; pharyngitis: 2315, 2172; bronchitis: 1515, 988; influenza: 515, 309; acute URI: 2071, 1349; and otitis media: 411, 185.

DIAGNOSIS-SPECIFIC PROFILES FOR JUDGING QUALITY OF CARE

5. A panel of physician judges in several primary care specialties, modes of practice, and metropolitan/nonmetropolitan areas of the country agreed at a high level about the appropriateness of the use or nonuse of numerous elements of care for these respiratory infections. Their judgments thus provided a valid basis upon which to construct age and diagnosis-specific profiles of quality of care.

6. These profiles were based, in effect, on explicit process criteria. They took account simultaneously of (a) the presence of elements of care considered by the physician panel to be required for adequate care, (b) the absence of unnecessary or contraindicated services (mainly drugs), (c) the presence of elements considered unnecessary or contraindicated, and (d) the absence of elements of care considered necessary. Positive quality-of-care scores would be produced by (a) and (b), negative scores by (c) and (d).

7. A number of services were found to be basically irrelevant to the quality-of-care profiles for some conditions. This was true, for example, for followup visits for all conditions except bronchitis and otitis media in children, and for non-narcotic analgesics and a variety of cough and cold preparations and decongestants in oral form. Essentially the judges considered both using and not using these services for these respiratory illnesses to be "probably acceptable."

8. The judges were nearly unanimous in rating certain services or medications as unacceptable for these conditions. These included all the symptomatic medications in injectable form, antibiotics in any form for influenza and acute URI, virtually all tetracycline except when used for bronchitis in adults, and all forms of lincomycin.

9. Three levels of satisfactory care--high, acceptable, and minimal--could be defined from the ratings provided by the physician panel for each of the age- and diagnosis-specific profiles. (The sum of these three levels is referred to as "adequate care" in this report.) A residual "unacceptable" category was also defined.

10. Empirically, only two levels of satisfactory care (high and acceptable), together with the unacceptable category, would have been needed to characterize episodes for the quality-of-care analyses. The minimal level was only rarely used.

QUALITY-OF-CARE ANALYSES

11. Physicians differed markedly in the quality of care they provided, judged by the percentage of episodes receiving adequate care (the sum of episodes scored minimal, acceptable, and high according to the diagnosis-specific profiles) and by the percentage of episodes scored high. In general, solo practice MDs provided better care than DOs, board-certified solo practice and group practice physicians better than their noncertified counterparts, and nonoutliers better than outliers. In some cases, outlier MDs provided better care than nonoutlier DOs. Generally, certified group practices, which were typically large, multi-specialty groups composed of certified MDs, delivered the best care.

Examples of the major findings about differences* among types of physicians include the following. These results refer only to strep throat, pharyngitis, acute bronchitis, influenza, and acute URI; see below for otitis media.

(a) For these five conditions in Period I, providers ranked as follows in the proportion of episodes to which they rendered adequate care (percentages in parentheses): certified group practices (70 percent); noncertified groups (62); certified MDs (55); noncertified MDs (42); certified DOs (38); and noncertified DOs (19). Thus, in Period I, both being certified and being in a group practice were factors predictive of better quality care.

In Period II, the rankings suggested that group practice was less important a factor and certification status was more important. Certified groups still had the best record (74 percent of the episodes had adequate care), but they were followed by certified MDs (68), noncertified MDs (57), noncertified groups (56), noncertified DOs (49), certified DOs (42).

(b) The expectation that, among solo practitioners who were not board certified, outlier physicians would give worse care than nonoutlier physicians was confirmed. Comparisons between like categories of MDs and DOs consistently favored the former. The percentages of episodes receiving adequate care in Period I were as follows for each category: nonoutlier MDs (56 percent); outlier MDs (21); nonoutlier DOs (24); outlier DOs (7). In Period II, the

*All these differences were statistically significant.

percentages were higher but the same patterns were observed: nonoutlier MDs (62 percent with adequate care); outlier MDs (48); nonoutlier DOs (55); and outlier DOs (31).

12. As implied by the findings just cited, care for these five conditions improved between Periods I and II. Of the 6807 episodes in Period I, 36 percent received adequate care, as did 55 percent of the 5291 episodes in Period II. High quality of care was recorded for 15 percent of Period I episodes and 23 percent of Period II episodes. These differences over time achieved statistical significance.

13. Significant improvements were observed for all five diagnoses. By Period II, the percentages of episodes of each diagnosis receiving adequate care were the following: strep throat, 45 percent; pharyngitis, 55 percent; bronchitis, 75 percent; influenza, 43 percent; and acute URI, 46 percent. Only for acute bronchitis could one conclude that care was likely to be reasonably satisfactory in this population.

14. Otitis media was judged, first, by a demanding profile that required the use of both appropriate antibiotics and a followup visit for adequate care and, second, by a more lenient profile that did not require a followup visit for a minimal (and thus adequate) score. Quality-of-care scores were fairly sensitive to which criteria set was used. Based on the more stringent criteria, certified groups and MDs provided better care than their noncertified counterparts, but the MDs gave better care than the groups.* Nonoutliers among the noncertified gave better care than the outliers.

*The numbers of otitis media episodes seen by some provider types (e.g., certified DOs) were too small for reliable analysis. Thus, statistical tests tended to be insignificant, and the otitis media results involving provider classifications are mainly impressionistic.

15. Care for otitis media improved over time, taking all providers together. Using the more stringent profile, 9 percent of episodes in Period I and 16 percent in Period II were judged adequate. Using the modified profile, the figures were 28 and 50 percent. In both cases, the improvements achieved statistical significance.

DISCUSSION

As noted in Chapter I and throughout the discussion of the quality-of-care analyses, this study extended a series of investigations of the differences in quality of care by type of physician and of the effect of the New Mexico EMCRO on the quality of ambulatory care delivered to a cohort of the Medicaid population. The substantive findings from this latest stage of the investigations did not differ in any meaningful way from earlier findings (see Lohr et al., 1980). Consequently, the results of this study relate more to methods for investigating entire episodes of care and assessing quality of care in settings where insurance claims are the main source of data than to any ongoing health policy debate. Some implications of these methodologic advances for the quality assessment field in general and for the PSRO program in particular are discussed below, together with some brief notes about the broader health policy issues to which these findings indirectly pertain.

EPISODES OF CARE

In the decade ahead, PSROs will probably take on more responsibility for reviewing care given in the outpatient setting (see, e.g., OTA, 1980), although priorities and budget constraints within the relevant agencies (especially the Health Care Financing Administration [HCFA]) will probably place ambulatory care review in line after improvements in inpatient review and expansion of long-term care review (see, e.g., Covell, 1980). Nonetheless, improved and augmented ambulatory care review must eventually make better use of both Medicaid and Medicare claims data, because such review, for practical reasons, cannot rely only on information in medical records, observational techniques, or surveys. Better ways of compiling and organizing such claims data may emerge from programs such as the National Medicaid Statistical System or the Medicaid Quality Control System (see, e.g., Galblum, nd), and from more sophisticated inputs from states or PSROS into the Medicaid Management Information System (MMIS) or the PSRO Management Information System (PMIS). These are all computer-dependent systems, suggesting that over the long run review techniques that apply to claims data and that can be easily adapted to computer applications (such as the "episodes-of-care" methodology used in this study) will be increasingly necessary and valued.

Using episodes of care as the unit of investigation will also be a more desirable approach to quality-of-care assessment than has been recognized or practical heretofore (Donabedian, 1978; Lohr et al., 1980). Episode techniques are needed if quality of care is to be evaluated taking several elements of care simultaneously into account. This situation would arise, for example, if a PSRO wished to investigate

the use of ancillary services (perhaps in an effort to identify physicians who seriously overuse diagnostic procedures) or to evaluate the use of various types of medications in cases when those medications should be employed only for confirmed diagnoses (thereby assuming the presence of some diagnostic procedure at or near the same time). These techniques are analogous in some ways to "profile analysis" and "focused review" as now being practiced within PSROs for inpatient care, but they have not received any meaningful applications in ambulatory care review to date.

Episode analysis is also more feasible now than in the past. For example, the simple computer algorithms used in the earlier EMCRO study provided valid episodes over three-quarters of the time; the more complex algorithms used here yielded valid episodes over nine-tenths of the time. Consequently, applying even a simple episode methodology to better organized claims data bases should prove quite fruitful for characterizing the ambulatory care delivered to various patient subgroups, for selected conditions, and by subgroups of providers in coming years.

Although the present episode algorithms were more detailed and conditional than the earlier ones, they were not substantially more complicated from the point of view of computer programming. Thus, in future years, review organizations, health services researchers, and other analysts should not find it much more difficult to use more complex episodes and computer techniques than to use simple ones. Adopting a more discriminating approach might be desirable on several counts. For one, it improves the face validity of the overall

methodology, by accounting more fully for variation in the natural history of these diseases and in individual patient response to them. For another, it permits investigation of a wider set of questions, such as whether any relationship seems to exist between disease complication rates and physician or patient characteristics.

PROFILES OF CARE

Composite profiles like those developed here can be a useful technique in quality assessment and assurance efforts. They reflect more accurately the range of decisions physicians need to make about treating patients. Typically, physicians are confronted with patients for whom a number of different decisions are needed: whether to do one or more laboratory tests, whether to begin antibiotics for an infection at the initial visit or after the results of cultures are known, whether to rely on patient cooperation and prescribe an oral medication instead of using an injectable one, and so on. Evaluating care by considering events one at a time--whether a test was done or followed up; whether a specific drug was or was not prescribed--is less satisfactory, conceptually at least, than evaluating events in a pattern (i.e., in the presence of all other pertinent elements).

Delineating a "quasi-continuum" of care may in many instances be preferable to making a simplistic dichotomous judgment that care is either good or bad. Profiles like those described in this report provide one basis for delineating such a continuum. They go beyond simple counts of services done or not done and take account of the simultaneous presence or absence of several (or many) factors important

to satisfactory medical care. They provide a mechanism for highlighting exemplary care, should that be considered of heuristic value.

Theoretically, "humanistic" and interpersonal elements of care might be incorporated into such profiles; the continuum might also be expanded by constructing, for example, more than one category of inappropriate care. The obstacles to such refinements are largely methodologic. Developing valid and reliable measures of the art of care, for instance, is a difficult but important task that is only now being explored in the quality-of-care field. Acquiring ratings from physician panels (like those in this study) about appropriateness of the use or nonuse of services for other acute conditions (not to mention chronic conditions) would also be a major undertaking.

In the meantime, the methodologic tools described in this study might be introduced into quality-of-care and utilization review efforts in areas outside the traditional PSRO concerns. For example, one subject of interest to federal agencies such as HCFA is federal reimbursement of allied health personnel and nurse practitioners for primary care services (e.g., "basic care services" in Medicare). A good deal of effort has been expended to evaluate "physician extender" practices along quality-of-care dimensions, but these efforts are hampered by the same difficulties that impede quality assessment of physician care. To the degree that profiles and/or episode analyses will provide more valid ways to review ambulatory care from physicians, they can be expected to improve assessments of paraprofessionals as well.

Similar points might be made about long-term care review, which is a growing concern of PSROs. The serious need for quality assessment and assurance efforts in skilled and intermediate nursing homes is unchallenged, but measurement capabilities and techniques are only poorly developed now (see Kane et al., 1979, 1980). Applications of these techniques to other settings and data bases should be tested. For example, an episode approach might be adapted to evaluate quality of care for exacerbations of chronic conditions or superimposed acute conditions in a setting where all care is channeled through either a single facility or a network of institutions.

Medical care problems of the elderly, in both the inpatient and outpatient settings, are clearly ones in which several elements of care should be evaluated simultaneously. Thus, efforts to define profiles that specify the concurrent presence and absence of good and bad features of care and allow for gradations in quality-of-care judgments should be a significant contribution to long-term care review. Some observers note that misuse of drugs is a (if not the) major deficiency of long-term care; profiles of care might be constructed in ways that would facilitate investigation of the proper use of injectable and oral medications for chronic conditions in elderly patients.

One advantage of profiles of this sort is that they are neither simple counts of individual services done or not done nor rates of services per some unit of time or per visit; they give a broader, more qualitative, assessment of care. Thus, they provide an easily interpreted way to characterize the practice of individual physicians or collection of physicians across very disparate conditions and patient populations. Using profiles that take a more holistic approach to

patient management may appear fairer and more realistic to providers and hence render the results of an evaluation more acceptable to those being evaluated. This "qualitative" feature may be especially pertinent if (or when) the national PSRO program is called on to draw comparisons among PSRO areas or subregions, among physicians of different specialties who see patients with similar conditions, or among different population subgroups.

For reasons noted in the discussions in previous chapters, no effort was made in this study to delineate more than one level of unacceptable care in these profiles. Nonetheless, attempting to define degrees or patterns of poor care may have some virtue. Some physicians, for instance, may give several different elements of poor care, suggesting that their overall strategies for patient management deviate substantially from common practice. Other providers, by contrast, may do only one thing poorly; for example, they may have a mistaken predilection for using powerful, broad-spectrum antibiotics for patients needing only a bacterium-specific drug.

Quality assurance strategies adopted by PSROs or others to improve quality of care might differ considerably, depending on which of these circumstances prevailed. Such strategies can include simply providing information, supporting or supplementing a broad range of continuing medical education (CME) efforts, or instituting financial sanctions. (Dropping providers from Medicaid or Medicare rolls is another option, but it implies a failure of the less harsh measures.) In ambulatory care review, PSROs need not and perhaps cannot pursue all such activities concurrently, and choosing initial interventions on behalf of individual

physicians or collections of providers might be guided in part by what patterns of unacceptable care are found. Thus, developing a capability to detect and examine such patterns would allow PSROs to tailor their surveillance activities and educational or sanction interventions much more efficiently than might otherwise be possible.

FUTURE DIRECTIONS OF PSRO AMBULATORY CARE REVIEW

As noted, the findings of this study pertaining to differences in quality of care differed in no material way from those of all the earlier studies. In general, among the solo practitioners, being an MD rather than a DO promoted better quality of care. Board-certified physicians in group or solo practice delivered better care than their noncertified counterparts. Finally, at least for certified physicians, practicing in a group apparently enhanced the quality of care rendered. The atypical nature of these groups (which were mainly large, multi-specialty, and academically oriented) suggests that this latter finding should be interpreted cautiously and warrants further study.

On the less positive side--which is, after all, where PSROs will need to concentrate their activities--two features can be highlighted. First, certain types of providers gave notably inferior care across the spectrum of diseases studied; in particular, DOs generally gave worse care than their MD colleagues and outliers among both MDs and DOs gave the poorest care of all. Second, the average level of care across all providers was not good enough to permit complacency.

Outliers or Typical Physicians as the Principal Concern?

The observation that these two types of problems (outliers giving very poor care and "typical" physicians giving less than satisfactory care) exist simultaneously may well be replicated across the nation. To deal adequately with them, PSROs will need to undertake two responsibilities in ambulatory care review more or less in tandem.

First, they will need to direct immediate attention to the levels of quality of care provided by physicians at the extreme poor end of the quality continuum. A PSRO could not long tolerate the levels of quality of care provided by outlier physicians and still retain its credibility; care given by the poorest providers should be improved even if efforts in that direction detract substantially from activities targeted on more accomplished physicians. Second, these analyses documented that all types of physicians exhibited deficiencies in handling the ordinary afflictions most often brought to primary care practices. One may reasonably infer that less prevalent conditions would not likely be handled at any better level of quality. Consequently, over the longer run PSROs will have to direct their attention to the care delivered by the typical physician as well.

One argument in favor of PSROs' devoting a substantial amount of attention to the outlier physician is that, by doing so, they can reduce the variance in the quality of care provided to these disadvantaged populations. This is of importance for ethical as well as practical reasons, because persons whose access to care is guaranteed by programs such as Medicaid typically have less freedom to choose physicians than do nondisadvantaged populations (regardless of the presumed intent of the programs to provide equal access to adequate care). Reducing the

variance in quality of care, therefore, by raising the performance of the poorest physicians without prompting any decline in the average level of care in the physician universe, is certainly a desirable outcome.

Improving the quality of ambulatory care in this manner (i.e., by concentrating on less competent physicians first) in effect raises the probability that persons eligible for federally funded care will eventually acquire broader access to better care. To the degree that PSRO review of care delivered through public programs has a spillover effect on care delivered through the private sector to non-Medicaid patients, it raises the probability that these persons, too, will benefit. Such efforts may thus contribute to a more just distribution of national health resources (all other things equal). For persons who reside at distances or in circumstances that preclude their seeking care from the best physicians--Medicaid families with little or no means of transportation from outlying areas to metropolitan sites or within urban areas, or poor and near-poor families not eligible for Medicaid--this is no mean achievement.

National or Local Criteria?

PSROs can base their evaluations on norms, criteria, and standards developed by physician consensus at the local level, although the National PSRO Council can reject such local standards if they deviate from "model" standards in ways that are not medically justified. Whether PSROs emphasize local or national quality-of-care standards will be an important question in coming years. Locally developed and

validated criteria do have some advantages. Not the least of these is that, according to implementation theory and experience in other fields (see, e.g., Williams, 1976), physicians may be more likely to accede to PSRO standards that they perceive are the locally accepted and customary ones than to standards handed down from centralized agencies. Thus, emphasizing criteria developed at the community level or even at the level of a PSRO area may facilitate implementation of ambulatory care review.

For the federal PSRO program, however, an emphasis on local criteria-setting has disadvantages. If each of the nearly 200 areawide PSROs tries to develop criteria from scratch, then setting criteria locally is likely to be very inefficient and costly. If community-derived criteria promote notably different levels of use of services in different localities, it may lead to situations in which health services are inequitably distributed to various disadvantaged populations (who are not in a position to move to different locations). Consequently, the preferred long-run policy is probably to try to develop nationally applicable criteria. At least as a beginning, such criteria might define "minimum" levels of acceptable care.

As noted earlier, PSROs will need to base much of their ambulatory care review efforts, at least for Medicaid, on insurance claims data. Developing "national" criteria applicable to such data will not be easy. Although a wide variety of criteria sets exist for numerous outpatient conditions, most of them, unfortunately, rely on data from medical records, not from insurance claim forms. The simple criteria developed for this study, and the criteria being developed by the Rand Health

Insurance Study (which is a nationwide longitudinal study) (Brook et al., 1980), relate directly to insurance claim data. Thus, they will be useful guides about the types of national standards that might be promulgated when and if insurance claims data are the principal source of information for peer review of ambulatory care.

Claims Review, Medical Record Review, or Both?

Despite the need to develop better methods of claim-based review, PSROs may in future years wish to adopt at least some aspects of office-based review, especially the performance review approach to ambulatory care review pioneered by the American Society of Internal Medicine (ASIM) (see, e.g., Farrington et al., 1980). The ASIM approach involves comparing performance as recorded in the office medical record with predetermined criteria for adequate quality of care. (These methods themselves could be adapted to include episodes of care, because defining entire episodes is easier with medical record data than with insurance claims data. Moreover, employing profiles of care as the evaluative criterion is compatible with the ASIM approach). As currently promulgated, performance review is strictly voluntary, locally conducted, and confined to a single specialty; it is not directly related to review of care given under the auspices of public programs.

Wholesale adoption of this performance review approach by the national PSRO program would not seem feasible for several reasons. First, it is determinedly local in emphasis, and as discussed above, this would not appear to be entirely compatible with the needs or goals of the national PSRO program. Second, it is voluntary in nature,

whereas the PSRO program is not (although participation in local PSRO review efforts is a voluntary decision by individual physicians). Third, the costs of a medical-record-review approach to ambulatory care, extrapolated nationally, would be high.

These drawbacks notwithstanding, some consideration might still be given to ways that joint or complementary efforts of PSROs and performance review programs might be facilitated. For instance, PSROs would continue to rely principally on reviewing claims data and invoking (where necessary) the educational or sanction interventions normally available to them; special interventions for physicians delivering markedly poor care might then be promoted through the voluntary performance review channels.

Costs and Quality of Care

Two related issues that this study did not address involved the costs of care and of peer review, but they are critical elements of an intense policy debate over the future configuration of health care delivery and financing. In the ambulatory care sector, we do not yet know much about the effects of peer review on costs. Although no one has conclusively shown that better quality of care is necessarily more expensive care, the most reasonable expectation is that, in the aggregate, as quality of care improves, costs will rise (to some unknown degree). The conventional wisdom has been that higher costs would come about from greater use of more extensively trained physicians, greater reliance on more technologically advanced services or tests, more use of needed services that were previously

underutilized, and possible substitution of unreviewed services for reviewed ones. To the extent that quality-of-care review is perceived to lead to such outcomes, it is assumed to raise health care expenditures; i.e., it not only may sustain the use of more and more expensive inputs to care, it adds the actual review costs to the total health care bill.

A few studies (e.g., Kane and his colleagues; the earlier New Mexico EMCRO evaluations; Goldberg and Jolly, 1980), however, lend partial support to two other views, namely, that good care may be no more costly than bad care, and that successful efforts can be mounted to curtail or eliminate use of services that have essentially a zero or even negative marginal product (as regards better health outcomes). Use of irrelevant ancillary services or excessively redundant use of relevant diagnostic tests and services can be reduced, for example, without prompting increases in use of different but equally unnecessary services. Less expensive but equally efficacious drugs may be substituted for more costly ones. In some settings, allied health professionals may be substituted for physicians with no loss in quality and with, perhaps, some cost savings.

PSRO quality-of-care review in the ambulatory sector will probably not "pay its own way"; it is certainly unlikely to realize net savings to the federal health budget (i.e., reductions in dollar expenditures over and above the program's costs). If such review can bring about even some of the improvements in medical care delivery just noted, however, one might reasonably argue that the positive benefits in the form of better care will balance, in some sense, the potentially higher

costs of that care. (Questions such as whether we, as a nation, wish to purchase better care in this fashion, how much we are willing to spend to acquire such care, and whether we will continue to commit such funds largely on behalf of our disadvantaged citizenry are ultimately social, political, and ethical in content, and go well beyond the present discussion.)

Growing public and professional attention to these cost/quality issues implies that comparing the costs and benefits of health care, or comparing costs among alternative patterns of care, will eventually be concerns even for the PSRO program. Cost/benefit and cost/effectiveness analysis techniques have received no real applications within PSROs to date (OTA, 1980), and much theoretical and applied research in these areas lies ahead. Nevertheless, given the importance that cost control has as a legislative and bureaucratic rationale for PSROs, these techniques will be indispensable components of quality-of-care efforts in this decade and beyond. In investigating these types of questions, episodes of care and comprehensive profiles of care may be helpful research techniques, especially if they allow for review of procedures and services for which costs may be high but the marginal benefits positive.

Some may even contend that PSRO ambulatory care review should be aimed primarily at containing medical care costs, and only secondarily at improving quality of care. Pursuing this monetary objective would be a departure from, if not a perversion of, the intent of ambulatory care review, at least as viewed by the medical community. Moreover, as discussed above, successfully controlling costs through ambulatory

review can be regarded as highly improbable--a judgment that is only reinforced by the PSRO experience with inpatient review.

Other developments in the health care arena, such as a shift away from centralized regulation and planning and toward pro-competitive, "market forces" delivery and financing schemes, may prove more effective, over the long run, in controlling health care costs. To the degree that such schemes may have an inherent tendency to force the use of services to levels below what is optimal (e.g., if providers were to compete partially on price and thus skimp on services), quality of care may become the dominant concern of PSROs in the years ahead. Thus, emphasis on quality-of-care review in the ambulatory care sector can be justified not only by the need to be sensitive to potentially detrimental tradeoffs between costs and quality, but also by the positive benefits achievable in the form of better medical care throughout the nation.

SUMMARY

This study demonstrated that Medicaid claims data could be used to create valid episodes of care for common respiratory illnesses. It showed that the computer programs to create those episodes could be sufficiently flexible to accommodate progressions of illnesses from less to more serious bacterial infections. The work substantiated the claim that quality of care could be judged on the basis of comprehensive profiles of care. These profiles specified both the presence and absence of, respectively, good and bad elements of care, and allowed for

judgments of high, acceptable, minimal, or unacceptable. They provided a "quasi-continuum" by which to differentiate the care rendered by different types of physicians and over time, and they allowed for judgments based on underuse as well as overuse of services. Finally, the study confirmed findings of earlier work that peer review undertaken by local physicians can improve the level of quality of care provided to disadvantaged persons.

Quality of care for these widely prevalent acute conditions was often below adequate levels: in some cases, no more than 50 percent of the episodes of colds and sore throats were treated satisfactorily. The deficiencies in quality of care varied according to certain professional characteristics of physicians. Typically, certified groups gave the best care of all; board-certified physicians (both group practices and MDs in solo practice) gave better care than those without such certification, MDs gave better care than DOs, and a few "outliers" accounted for much of the poor care observed, especially among the noncertified physicians.

The New Mexico peer review organization brought about substantial improvements in the care to the Medicaid population, by curtailing the use of inappropriate injectable drugs (particularly antibiotics) without inducing any decrements in care (such as increased inappropriate use of oral drugs). Because the greatest improvements in care were observed among physicians who had had the poorest records initially, it was concluded that the overall variance in quality of care to this disadvantaged population had been reduced, a finding that was considered important from the clinical, ethical, and policy points of view. .

APPENDICES

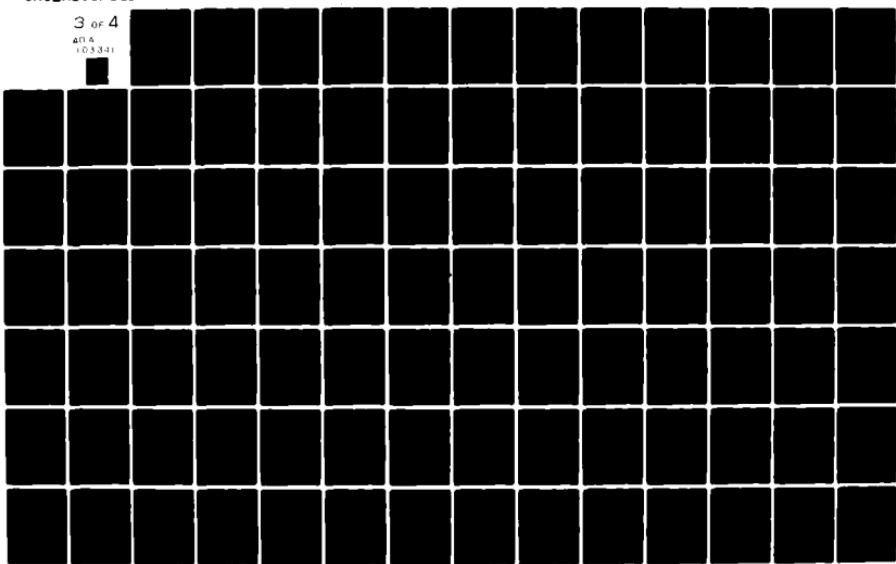
PRECEDING PAGE BLANK-NOT FILMED

AD-A103 341 RAND CORP SANTA MONICA CA
QUALITY OF CARE IN EPISODES OF COMMON RESPIRATORY INFECTIONS IN--ETC(U)
OCT 80 K N LOHR
UNCLASSIFIED RAND/P-6570

F/G 6/5

NL

3 of 4
404
103-341



APPENDIX A
VALIDATION OF EPISODE RULES

RAND HEALTH INSURANCE STUDY DATA

The types of data that can be recorded on the HIS Physician Outpatient claim include physician visits (to offices, emergency rooms, and homes), injections administered, laboratory tests and other diagnostic procedures performed, outpatient services rendered (e.g., minor surgery), oral medications for which a prescription is written and given to the patient, and prescriptions filled by the physician.

Another element of this claim form is a "treatment history code," which allows physicians to link services into episodes by designations such as "repeat acute care," "chronic routine care," "chronic flareup, initial care," and "chronic flareup, repeat care." Even if, for example, the physician changes the diagnosis from one condition to another (for example, from acute URI to bronchitis), he or she may still indicate one of the repeat visit codes, suggesting that some underlying condition is ongoing.

With most claims data sets, prescription drugs are typically billed on a claim form submitted directly from a pharmacy, and such a claim often lacks information such as prescribing physician, diagnosis for which the medication was prescribed, and so forth.

These difficulties are obviated to some extent in the HIS by two features. First, physicians who dispense medications can bill for them on the Physician Outpatient claim. Thus, date of service will match,

other relevant information such as dosage, generic and brand names, etc., is available, and, most important, the prescription will be related directly to the diagnosis for which it was given.

Moreover, prescriptions given to patients to be filled elsewhere must still be recorded by the prescribing physician, with diagnosis, or the Physician Outpatient claim. If a HIS Pharmacy claim for the same drug has the same date of service or a date reasonably close in time (either before or after), it can plausibly be linked directly to the episode of care. Medications from both the Physician Outpatient and Pharmacy claims are coded in the HIS with information such as therapeutic code, generic and/or brand names of drugs, dosage, dates of service, and so forth. Thus, matches between prescription noted on the Outpatient claim and those filled by a Pharmacy claim are considered quite reliable.

Two exceptions to the preceding paragraph should be noted. First, refills of drugs that appear only on Pharmacy claims may be quite remote in time from the Physician Outpatient claim (if any) to which they might be linked, making any direct match difficult. Similarly, prescriptions that a physician telephones to a pharmacy may not be recorded on any Physician Outpatient claim, meaning that no diagnostic data may be available by which to link the prescription directly to an episode. In the HIS data set used for the validation studies, however, prescriptions lacking any diagnosis made up less than 20 percent of all prescriptions; of those without diagnoses, many were for drugs that are typically prescribed on a long-term basis or for chronic conditions (e.g., tranquilizers, antihypertensives, insulin) and others were obviously refills of prescriptions for which diagnoses were already available.

FIGURE A.1

00122278 PATIENT/PROFILES GENERATED FROM DAYTON CLAIM

N0251662 AGE: 20 SEX: FEMALE PLANS:

INVENTORY FILE

PAGE 32

	DIAGNOSIS	ENGLISH TRANSL.	RUSSIAN TRANSL.	ENGLISH TRANSL.	RUSSIAN TRANSL.
	DIAGNOSIS	ENGLISH TRANSL.	RUSSIAN TRANSL.	ENGLISH TRANSL.	RUSSIAN TRANSL.

75/06/19 27563

0 A1: 90040 BRIEF OFFICE VISIT, ESTAB PT 775.7 LYMPHADENOPATHY NOS
 A2: RFV 4100 9400
 A1: 05022 COMPLETE BLOOD COUNT 775.7 LYMPHADENOPATHY NOS
 01: 89 TETANUS TOK 781.2 FOR TETANUS TOKIO

75/08/13 27418

3 A1: 90050 LIA OFFICE VISIT, ESTAB PT 785.4 VAGINAL DISCHARGE NEC
 A2: RFV 9040
 A1: 00150 PAPANICOLAOU SMEAR 785.4 VAGINAL DISCHARGE NEC
 C1: 37 MORINYL 1 PLUS 50

75/11/10 27715

6 A1: 90040 BRIEF OFFICE VISIT, ESTAB PT 777.6 THROAT PAIN
 A2: RFV 3120 5200
 A1: 07060 CULTURE, BACTERIAL 777.6 THROAT PAIN
 C1: 49 V-CILLIN K 777.6 THROAT PAIN
 C1: 43 TUSSENO EXP. 777.6 THROAT PAIN
 01: 0020W07 V-CILLIN K
 62: 250 MG Q1D, 3-4X0 1/2-1 CAP, TAB
 63: Q6H, Q1D, 3-4X0 1/2-1 CAP, TAB

75/11/29 27726

6 A1: 90040 BRIEF OFFICE VISIT, ESTAB PT 774.6 ACUTE PHARYNGITIS
 A2: RFV 3200
 C1: 93 TETRACYCLINE 774.6 ACUTE PHARYNGITIS
 61: 0354875 ACHROMYCIN (CAP.)
 62: 250 MG 1/2-1 CAP, TAB
 63: Q6H, Q1D, 3-4X0 1/2-1 CAP, TAB

76/04/14 27863

A1: 64592 SYPHILIS TESTS

V00-1 VD SCREENING PROGRAM

Figure A.1 gives an example of an HIS patient profile, showing all the information coded from claims submitted on 5 different days for a variety of services and conditions for a 20-year old woman. For the purposes of the validation work, the information about services for throat pain and acute pharyngitis (on 11/18/75 and 11/29/75, respectively) is of interest. The data associated with throat pain, for example, show that the patient had a brief office visit with a physician who regarded her as an established patient ("A1" in the far left column), that during the office visit she had a bacterial culture of the throat (also "A1"), that the physician prescribed an antibiotic (V-cillin K) and an antitussive/expectorant (Tussend) for the throat pain ("C1"), and that the patient had the prescription for the antibiotic filled at a pharmacy on the same day as the office visit ("G1"). From the patient profile, the Pharmacy claim prescription can be linked to the diagnosis of the office visit (and hence of the episode) on the basis of the date of service, the name of the drug, and the therapeutic code (in this case, 48).

In the HIS profiles, diagnoses are coded by HICDA-II codes and all symptoms by a modified and expanded version of the National Ambulatory Medical Care Survey Symptom Classification (NAMCS). All services are coded by the 1974 California Relative Value Scale (CRVS) codes, and injectable and oral medications by the AMA Drug Evaluation code (second edition). The notations such as "A1" are HIS identifiers for the type of claim or service represented by the entry. The patient profile is not ordered systematically by type of insurance plan to which the

FIGURE A.2

EPISODES FROM THE HEALTH INSURANCE STUDY

DIAGNOSIS/ICD NO: Pharyngitis/tonsillitis
(H.6) Strept throat

(H.6) Stack the root

Throat signs

190

Number	Dates and MCDA	For Diagnosis: Same Visit or Related Visit				For Other Diagnosis: Same Visit or Related			
		Visit	Lab Test	Drugs		Visit	Lab Test	Drugs	
				Injections Given	Oral Pres. Filled			Injectors Given	Oral Pres. Filled
1. 250223	6/15/76 (463)	Visit		Therapeutic Name Code	Therapeutic Name Code	35	Decadron	53 44 27	V. bromycin Tessalon Second
2. 251638	7/14/75 (462)	Visit	Culture					53 44 27	V. bromycin Tessalon Second
3. 251654	8/15/76 (776)	Visit						43	Act. Cd
4. 251662	11/18/75 (776)	Visit	Culture					42	Nasal Spray
5. 251725	11/29/76 (462)	"	"					48 43 48 53 53	V-cillin K Tussend V-cillin K Tetracycline Achromycin
								43 53	Comonol Tetracycline
									46 Ornade
									6.7 My qrotan

person's family had been assigned or by any other demographic characteristics; it is ordered within each person by date of service.

ABSTRACTING HIS EPISODES

Figure A.2 displays the special form used for recording HIS data into episodes, with the relevant columns noted in parentheses: (1) an ID number for the episode and the patient ID number (column 1); (2) dates of service and diagnostic codes (column 2); (3) services for which the physician specified at least one of the above four diagnoses at either the initial visit or a subsequent visit within that episode (columns 3-6 headed "For diagnosis: Same visit or related visit"); these services included physician visits, laboratory and diagnostic tests, injections administered (with therapeutic code and name of drug), prescriptions for oral medications written for the patient, and oral medications prescriptions actually dispensed (with therapeutic code and name of drug); (4) Services for which the physician specified a diagnosis other than one of the four sore throat diagnoses at either the initial visit or a subsequent visit (columns 7-12); these could include laboratory tests or medications given for a chronic condition at the same visit as for the acute diagnosis under study; (5) Pharmacy claim prescriptions that could not be linked to a diagnosis in or around the time of the episode itself or would not be considered relevant for the diagnoses being studied (such as birth control pills or tranquilizers (again, columns 7-12).

Not recorded were the following: (1) Certain medications data (form and dosage); (2) CRVS or NAMCS codes; (3) patient-related

information (e.g., age, sex, insurance plan); and (4) in some cases second, third, or fourth diagnoses completely unrelated to sore throat or respiratory system disease (e.g., obesity) for which, evidently, nothing was done.

In Figure A.2, drugs enclosed in a box were written but not necessarily filled; that is, they appeared as a "C1" code on the patient profile (see Figure A.1). Drugs not enclosed in a box were either provided by the physician (in which case a diagnosis is associated with them directly) or filled by a pharmacy.

The medications, diagnostic tests, and procedures used in these analyses were restricted to the following:

Medications: antibiotics; antitussives and expectorants; antihistamines; antinauseants/antiemetics; narcotic and nonnarcotic analgesics; various nasal, cough, and cold preparations and decongestants, and bronchial dilators.

Tests and procedures: chest x-rays; sinus x-rays; common blood tests (CBC, WBC, heterophile, hematocrit and hemoglobin, etc.); throat or strep cultures; cultures from other sources such as ear, sinus, etc.; urinalysis and urine cultures; and various multi-item "panel" or "profile" tests.

Excluded from consideration, therefore, were medications such as steroids and hormones, chronic disease medications, vitamin preparations, and allergy and immunization injections, and procedures or tests such as Pap smears or glucose testing. These exclusions were planned for the quality-of-care assessments using the New Mexico EMCRO data, and were used here to preserve comparability.

Many more services, however, were recorded on the abstracting form than would be included in the final respiratory episodes, precisely so that errors from the "New Mexico" rules could be detected. Services that could even remotely be considered related to "sore throat" and that occurred as much as two months after a visit for one of the sore throat episodes were recorded initially as a single "episode."

Examples of what were and were not included in an HIS episode can be seen from Figure A.2. Episode No. 4 has the following characteristics: a physician visit, throat culture, and two prescriptions written (one for an antibiotic [V-cillin K] and the other for an antitussive [Tussend]) on 11/18/75, all for the diagnosis sore throat (HICDA code 777.6); one prescription filled (for V-cillin K) on 11/18/75; a visit and a prescription written for a different antibiotic (tetracycline) on 11/29/76, all for the diagnosis pharyngitis (HICDA 462) on 11/29/79.* Episode No. 5 has the following characteristics: a physician visit two injections (one for an expectorant (Gomenol) and the other for an antibiotic (tetracycline)), and a prescription written for a cold preparation (Ornade), all for the diagnosis tonsillitis (HICDA 463). The prescription was not filled; i.e., in this instance no Pharmacy claim for this prescription was filed within six months of the visit. (The prescription for the antihypertensive (Hygroton) was not included in the episode, but was recorded for the sake of completeness.)

*As noted earlier, for the main analyses in the validation work, these two diagnoses are considered equivalent, although for other analyses, such as quality-of-care assessment, they would probably need to be distinguished.

APPENDIX B
METHODOLOGIC DETAILS OF DESIGNING
THE QUALITY-OF-CARE PROFILES

METHODS

QUESTIONNAIRES

For data entry purposes only, the four quality categories of the questionnaires were coded as follows: high, "1"; probably acceptable, "2"; probably not acceptable, "3"; and poor, "4." No connotation as to an interval/ratio level of data was intended to be implied by such scores.

Missing data and ostensible errors were rare. Three judges omitted several contiguous variables for one or another disease/age category; their responses for those few items were obtained by telephone. "Errors" (e.g., seemingly irrational or careless responses) by the judges were also very rare; these include rating a service or medication as "poor" care when what undoubtedly was intended was "high" (as evidenced by the pattern of responses for similar services or medications). For example, one judge marked "poor" care for "no heterophile or other blood test" in acute URI among children 0 to 7 years; given that judge's pattern of responses about all the other tests in that disease/age group, however, the "high" or "probably acceptable" notation was no doubt intended. Mistakes of this nature, which were probably caused by the relatively long and complex rating task, were so infrequent that they had no effect on the overall ratings or

quality-of-care Profiles, and they were simply ignored. (They did, however, lower somewhat the overall level of agreement among judges.)

MEDIAN RATINGS

Tables B.1 to B.12 display the frequency distributions of ratings for each age/diagnosis category. The median "consensus" ratings were determined manually from these distribution.

The median was used as a guide to the composite rating for two reasons. First, these ratings are considered ordinal, not interval in nature. As a measure of central tendency for data of this type, the median may be more appropriate than, say, the mean, in part because it is less sensitive to extreme values (e.g., to one judge's having mistakenly rated a service high when all others rated it probably not acceptable and poor). Second, attempting to calculate a mean implies that a numerical score must be assigned to each category (such as "1" equals high ... "4" equals poor). This in turn implies that some assumption must be made as to what numerical interval the score is believed to represent. For example, an assumption has to be made as to whether "1" is the midpoint of an interval from 0.5 to 1.4, or is an interval from 1.0 to 1.9. Although some assumptions also must be made in using a median (e.g., what to do in case of ties), they may be less open to questionable interpretation than those for the mean in situations involving categorical data.

Situations may arise in which no "true" median category can be computed. A 50:50 split between two quality-of-care categories may occur, for instance. "Ties" of this sort were relatively rare; less

than 5 percent of the ratings involved a tie. They were dealt with by always taking the higher rating as the median, as a means of giving better score in otherwise ambiguous cases.

One drawback to median ratings is that they do not by themselves reflect the nature or extent of divergence or disagreement among judges when such information might be of interest. For example, certain medications or services were unanimously considered poor care (see, e.g., injectable narcotic analgesics for several conditions), whereas other medications or services received poor ratings from only one more than half of all judges (see, e.g., bronchial dilators in pharyngitis). Both, however, were accorded a median rating of "poor." One should probably place less confidence in the latter median rating than in the former and take account of the range of ratings when using the median for setting criteria.

This problem (that the median necessarily masks some information about the extent of agreement) occurred chiefly for services rated probably acceptable. Some services/medications with a median rating of probably acceptable were considerably more "acceptable" than others, in the sense that their range of ratings was higher (e.g., perhaps one judge with high, eight with probably acceptable, and one with probably not acceptable, compared with five judges endorsing probably acceptable and five probably not acceptable). In defining the Profiles, services with the "better" probably acceptable rating were placed in the Acceptable classification in the Profile and the services with the "worse" rating were placed in the Minimal classification.

A related disadvantage with the median is that it can be a category that relatively few judges endorse. For example, one distribution of scores that could be observed is as follows: probably acceptable, four judges; probably not acceptable, one judge; and poor, five judges. The median in this instance is the probably not acceptable category (given the rule about breaking "ties"). This phenomenon was observed occasionally in the 0-to-7 age group or when missing data led to an even number of judges in the 8-and-older group.

This disadvantage with using the median ratings for the Profiles is not critical. It occurred only with the probably acceptable and probably not acceptable ratings. All probably acceptable services/medications were allowed at least for Minimal quality of care in the Profiles, as were all probably not acceptable services/medications except antibiotics. Only ten antibiotics received a probably not acceptable rating in a situation in which fewer judges marked probably not acceptable than any other category (i.e., in which probably not acceptable was the category with the minority of endorsements).

Placing these drugs in the Profiles was decided on a case-by-case basis. Five cases were oral cephalosporin, which recent literature suggests is unlikely to be a suitable drug for these conditions (at least in the forms available in the period covered by this study). Four of the remaining five involved various combinations of ampicillin and penicillin; in some cases, these were upgraded to probably acceptable to make them consistent with the ratings for other combinations of the same two drugs for the particular disease/age category. The last involved oral sulfa alone and was left as probably not acceptable.

AGREEMENT AMONG JUDGES

Several approaches to examining levels of agreement among judges were considered, including inter-rater reliability coefficients (especially kappa and weighted kappa), rank order coefficients, product moment coefficients, and simple proportion of agreement.

The more complex techniques were rejected in favor of the simple proportion of agreements either because they would give odd results in situations involving numerous ties (e.g., the rank correlation and product moment methods) or because they might give misleadingly low coefficients when the level of agreement among judges was actually extremely high and the marginals thus skewed in one direction (e.g., kappa). Even more elaborate measures involving weighting for observed tendencies of judges to choose higher or lower scores were rejected on the grounds that the additional computations and analysis would probably not provide sufficient information over and above that conveyed by the simple proportion of agreement, especially with such small numbers.

As noted in the Methods section of Chapter III, two definitions of agreement were used to examine the level of agreement among judges: "exact" and "dichotomous."

To give some perspective to "exact" agreement, assume that 10 judges must rate items into one of four categories. (This is the situation for the 0-to-7 age group.) If as few as 2 judges disagree with each other and with their colleagues about the ratings for all 69 items (not necessarily the same 2 judges in all cases), the level of exact agreement will be only 62 percent. If 3 judges disagree with

their colleagues and each other, the level will be 47 percent. The "worst case" (when 3 judges endorse one category, 3 judges endorse a second, and 2 judges each endorse the third and fourth categories) produces a level of exact agreement of only 18 percent. This is the reasoning that leads one to interpret exact agreement as fairly stringent. With 11 judges, the percentages are a bit different but the pattern is analogous. Following the above examples, 2 judges disagreeing produces 65 percent exact agreement, 3 judges disagreeing produces 51 percent agreement, and the "worst case" is only 16 percent agreement.

To see why dichotomous agreement is so much less rigorous, assume the same examples as above. Two judges disagreeing produces either 80 percent or 64 percent agreement (depending on whether they themselves split along the satisfactory/unsatisfactory dimensions), 3 judges disagreeing produces 64 percent agreement, and the "worst case" gives 47 percent agreement. With 11 judges, these examples would be either 82 or 67 percent for 2 judges, 67 percent for 3, and 45 percent for the "worst case."

200
 Table B.1
STREP THROAT 0-7 years

<u>TYPE OF DRUG OR SERVICE</u>	<u>NUMBER OF JUDGES RATING EACH SERVICE AS:</u>					<u>MEDIAN CATEGORY^c</u>
	<u>High</u>	<u>PA^a</u>	<u>PNA^a</u>	<u>Poor</u>	<u>Total^b</u>	
Antibiotics						
None				10	10	Poor
IM SA Penicillin			3	7	10	Poor
IM LA Penicillin	9	1			10	High
IM Ampicillin		3	3	4	10	PNA
IM Lincomycin			1	9	10	Poor
IM Tetracycline			1	9	10	Poor
Oral Penicillin	10				10	High
Oral Ampicillin	1	4	4	1	10	PA
Oral Lincomycin			1	9	10	Poor
Oral Tetracycline			1	9	10	Poor
Oral Erythromycin	6	3	1		10	High
Oral Cephalosporin		4	2	4	10	PNA
Oral Sulfa			4	6	10	Poor
IM SA Pen + Oral Pen	4	5		1	10	PA
IM SA Pen + Oral Amp	1	3	5	1	10	PNA
IM SA Pen + Oral Tetra			1	9	10	Poor
IM SA Pen + Oral Erythrom	2	4	2	2	10	PA
IM SA Pen + Oral Lincom			1	9	10	Poor
IM LA Pen + Oral Pen	2	3	3	2	10	PA
IM LA Pen + Oral Amp		2	6	2	10	PNA
IM LA Pen + Oral Erythrom	3		5	2	10	PNA
IM LA Pen + Oral Lincom			2	8	10	Poor
IM LA Pen + Oral Tetra			3	7	10	Poor
IM LA Pen + Oral Sulfa			4	6	10	Poor
IM Amp + Oral Pen	4		3	3	10	PNA
IM Amp + Oral Amp	2	5	3		10	PNA
IM Lincom + Oral Pen			2	8	10	Poor
IM Lincom + Oral Amp			2	8	10	Poor
IM Lincom + Oral Tetra			1	9	10	Poor
IM Lincom + Oral Erythrom			1	8	9 ^d	Poor
IM Lincom + Oral Lincom			1	9	10	Poor
IM Tetra + Oral Tetra				9	9 ^d	Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.1, cont.
STREP THROAT 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives				10	10	Poor
Oral Antitussives	1	5	2	2	10	PA
No Antitussives	6	4			10	High
IM Antihistamines				10	10	Poor
Oral Antihistamines		4	2	4	10	PNA
No Antihistamines	7	3			10	High
IM Antinauseants				9	10	Poor
Oral Antinauseants		2	2	6	10	Poor
No Antinauseants	6	2		2	10	High
IM Narcotic Analgesics				10	10	Poor
Oral Narcotic Analgesics			2	8	10	Poor
No Narcotic Analgesics	7	1		2	10	High
IM Non-narcotic Analgesics				10	10	Poor
Oral Non-narcotic Analgesics	1	7	1	1	10	PA
No Non-narcotic Analgesics	5	4	1		10	High
Oral Decongestants		6	2	2	10	PA
No Decongestants	6	3	1		10	High
Bronchial Dilators		2	3	5	10	PNA
No Bronchial Dilators	6	3		1	10	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

202
 Table B.1, cont.
STREP THROAT 0-7 years

<u>TYPE OF DRUG OR SERVICE</u>	<u>NUMBER OF JUDGES RATING EACH SERVICE AS:</u>					<u>MEDIAN CATEGORY^c</u>
	<u>High</u>	<u>PA^a</u>	<u>PNA^a</u>	<u>Poor</u>	<u>Total^b</u>	
Antibiotics						
Chest X-ray		2	6	2	10	PNA
No Chest X-ray	7	3			10	High
Sinus X-ray		1	7	2	10	PNA
No Sinus X-ray	8	2			10	High
WBC/CBC		7	3		10	PA
No WBC/CBC	5	4	1		10	High
Heterophile		6	3	1	10	PA
No Heterophile	6	4			10	High
Urinalysis		3	6	1	10	PNA
No Urinalysis	6	3	1		10	High
Strep/Throat Culture	9	1			10	High
No Strep/Throat Culture	1	2	4	3	10	PNA
Other Culture	1	2	6	1	10	PNA
No Other Culture	7	3			10	High
Panel/Profile		1	5	4	10	PNA
No Panel/Profile	7	3			10	High
Followup Visits	3	6	1		10	PA
No Followup Visits	1	5	4		10	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.2

STREP THROAT 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None				1	9	10 ^e
IM SA Penicillin			4	7	11	Poor
IM LA Penicillin	10	1			11	High
IM Ampicillin		3	3	5	11	PNA
IM Lincomycin			2	9	11	Poor
IM Tetracycline			2	9	11	Poor
Oral Penicillin	11				11	High
Oral Ampicillin		5	4	2	11	PNA
Oral Lincomycin			2	9	11	Poor
Oral Tetracycline		1	3	7	11	Poor
Oral Erythromycin	7	3	1		11	High
Oral Cephalosporin		5	2	4	11	PNA
Oral Sulfa			4	7	11	Poor
IM SA Pen + Oral Pen	5	5	1		11	PA
IM SA Pen + Oral Amp	1	4	4	2	11	PNA
IM SA Pen + Oral Tetra	2		2	7	11	Poor
IM SA Pen + Oral Erythrom	3	5	1	2	11	PA
IM SA Pen + Oral Lincom			1	10	11	Poor
IM LA Pen + Oral Pen	2	5	2	2	11	PA
IM LA Pen + Oral Amp		4	5	2	11	PNA
IM LA Pen + Oral Erythrom		5	4	2	11	PNA
IM LA Pen + Oral Lincom			3	8	11	Poor
IM LA Pen + Oral Tetra	1	4		6	11	Poor
IM LA Pen + Oral Sulfa			5	6	11	Poor
IM Amp + Oral Pen	4		3	4	11	PNA
IM Amp + Oral Amp	2		5	4	11	PNA
IM Lincom + Oral Pen	1		2	8	11	Poor
IM Lincom + Oral Amp			3	8	11	Poor
IM Lincom + Oral Tetra			2	9	11	Poor
IM Lincom + Oral Erythrom	1		1	8	10 ^e	Poor
IM Lincom + Oral Lincom			2	9	11	Poor
IM Tetra + Oral Tetra			1	9	10 ^e	Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.2, cont.

STREP THROAT 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives		1		10	11	Poor
Oral Antitussives	1	5	2	3	11	PA
No Antitussives	7	3	1		11	High
IM Antihistamines			1	9	10 ^e	Poor
Oral Antihistamines		3	3	4	10 ^e	PNA
No Antihistamines	8	2		1	11	High
IM Antinauseants		1	1	8	10 ^e	Poor
Oral Antinauseants		2	3	5	10 ^e	PNA
No Antinauseants	7	2		2	11	High
IM Narcotic Analgesics				10	10 ^e	Poor
Oral Narcotic Analgesics		1	2	7	10 ^e	Poor
No Narcotic Analgesics	8	1		2	11	High
IM Non-narcotic Analgesics			1	10	11	Poor
Oral Non-narcotic Analgesics	1	8	1	1	11	PA
No Non-narcotic Analgesics	6	4	1		11	High
Oral Decongestants	1	6	2	2	11	PA
No Decongestants	7	3		1	11	High
Bronchial Dilators		2	3	6	11	Poor
No Bronchial Dilators	7	2		2	11	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.2, cont.

STREP THROAT 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray		1	7	3	11	PNA
No Chest X-ray	9	2			11	High
Sinus X-ray		1	8	2	11	PNA
No Sinus X-ray	8	3			11	High
WBC/CBC		6	5		11	PA
No WBC/CBC	6	3	2		11	High
Heterophile		7	4		11	PA
No Heterophile	6	4	1		11	High
Urinalysis		2	8	1	11	PNA
No Urinalysis	7	3	1		11	High
Strep/Throat Culture		8	3		11	High
No Strep/Throat Culture	1	4	3	3	11	PNA
Other Culture		1	3	6	11	PNA
No Other Culture	8	3			11	High
Panel/Profile		1	6	4	11	PNA
No Panel/Profile	8	3			11	High
Followup Visits		3	6	2	11	PA
No Followup Visits	3	4	4		11	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

206
 Table B.3
OTITIS MEDIA 0-4 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None		1	1	8	10	Poor
IM SA Penicillin	1		1	8	10	Poor
IM LA Penicillin		2	2	6	10	Poor
IM Ampicillin	3	2	1	4	10	PA
IM Lincomycin			1	9	10	Poor
IM Tetracycline			1	9	10	Poor
Oral Penicillin	1	3	2	4	10	PNA
Oral Ampicillin	10				10	High
Oral Lincomycin			2	8	10	Poor
Oral Tetracycline		1		9	10	Poor
Oral Erythromycin	6	2	2	2	10	PA
Oral Cephalosporin		1	2	7	10	Poor
Oral Sulfa	1	3	2	4	10	PNA
IM SA Pen + Oral Pen	1	3	2	4	10	PNA
IM SA Pen + Oral Amp	3	5	1	1	10	PA
IM SA Pen + Oral Tetra		1		9	10	Poor
IM SA Pen + Oral Erythrom	3	3	3	4	10	PNA
IM SA Pen + Oral Lincom			2	8	10	Poor
IM LA Pen + Oral Pen		4	1	5	10	PNA
IM LA Pen + Oral Amp	2	5	3	2	10	PA
IM LA Pen + Oral Erythrom	3	3	3	4	10	PNA
IM LA Pen + Oral Lincom			2	8	10	Poor
IM LA Pen + Oral Tetra		1		9	10	Poor
IM LA Pen + Oral Sulfa	1	3	2	4	10	PNA
IM Amp + Oral Pen	1	4	1	4	10	PA
IM Amp + Oral Amp	5	4		1	10	High
IM Lincom + Oral Pen			2	8	10	Poor
IM Lincom + Oral Amp			1	8	9 ^e	Poor
IM Lincom + Oral Tetra			1	9	10	Poor
IM Lincom + Oral Erythrom			1	9	10	Poor
IM Lincom + Oral Lincom			2	8	10	Poor
IM Tetra + Oral Tetra		1		9	10	Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.3, cont.
OTITIS MEDIA 0-4 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives				10	10	Poor
Oral Antitussives	6	1	3	10	PA	
No Antitussives	7	3		10	High	
IM Antihistamines			1	9	10	Poor
Oral Antihistamines	1	8	1	10	PA	
No Antihistamines	4	6		10	PA	
IM Antinauseants			1	9	10	Poor
Oral Antinauseants	5	1	4	10	PA	
No Antinauseants	6	4		10	High	
IM Narcotic Analgesics	1	1	8	10	Poor	
Oral Narcotic Analgesics	2	3	5	10	PNA	
No Narcotic Analgesics	8	2		10	High	
IM Non-narcotic Analgesics	1	1	8	10	Poor	
Oral Non-narcotic Analgesics	5	5		10	High	
No Non-narcotic Analgesics	4	5		9 ^e	PA	
Oral Decongestants	2	7	1	10	PA	
No Decongestants	3	4	2	10	PA	
Bronchial Dilators	2	2	6	10	Poor	
No Bronchial Dilators	7	2	1	10	High	

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.3, cont.

OTITIS MEDIA 0-4 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray		1	5	4	10	PNA
No Chest X-ray	7	3			10	High
Sinus X-ray		2	7	1	10	PNA
No Sinus X-ray	6	4			10	High
WBC/CBC		6	4		10	PA
No WBC/CBC	4	6			10	PA
Heterophile		1	7	2	10	PNA
No Heterophile	4	6			10	PA
Urinalysis		2	6	2	10	PNA
No Urinalysis	4	6			10	PA
Strep/Throat Culture		7	2	1	10	PA
No Strep/Throat Culture	4	6			10	PA
Other Culture	2	7	1		10	PA
No Other Culture	2	7	1		10	PA
Panel/Profile		1	5	4	10	PNA
No Panel/Profile	5	5			10	High
Followup Visits	9	1			10	High
No Followup Visits			3	7	10	Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.4
OTITIS MEDIA 5-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None		1	1	8	10	Poor
IM SA Penicillin	1		4	5	10	PNA
IM LA Penicillin	2	4	1	3	10	PA
IM Ampicillin	2	1	2	5	10	PNA
IM Lincomycin			1	9	10	Poor
IM Tetracycline			1	9	10	Poor
Oral Penicillin	8	1		1	10	High
Oral Ampicillin	6	3	1		10	High
Oral Lincomycin			2	8	10	Poor
Oral Tetracycline		1	1	8	10	Poor
Oral Erythromycin	3	5	2		10	PA
Oral Cephalosporin		1	3	6	10	Poor
Oral Sulfa		4	2	4	10	PNA
IM SA Pen + Oral Pen	5	3		2	10	High
IM SA Pen + Oral Amp	2	4	2	2	10	PA
IM SA Pen + Oral Tetra		1	1	8	10	Poor
IM SA Pen + Oral Erythrom		5	2	3	10	PA
IM SA Pen + Oral Lincom		1	1	8	10	Poor
IM LA Pen + Oral Pen	3	4		3	10	PA
IM LA Pen + Oral Amp	1	5	1	3	10	PA
IM LA Pen + Oral Erythrom	1	4	2	3	10	PA
IM LA Pen + Oral Lincom			2	8	10	Poor
IM LA Pen + Oral Tetra		1	1	8	10	Poor
IM LA Pen + Oral Sulfa		3	1	6	10	Poor
IM Amp + Oral Pen	1	6		3	10	PA
IM Amp + Oral Amp	3	5	1	1	10	PA
IM Lincom + Oral Pen			2	8	10	Poor
IM Lincom + Oral Amp			1	8	9 ^e	Poor
IM Lincom + Oral Tetra			1	9	10	Poor
IM Lincom + Oral Erythrom			1	9	10	Poor
IM Lincom + Oral Lincom			1	8	9 ^e	Poor
IM Tetra + Oral Tetra			1	9	10	Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.4, cont.

OTITIS MEDIA 5-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives			1	9	10	Poor
Oral Antitussives	7			3	10	PA
No Antitussives	7	3			10	High
IM Antihistamines			1	9	10	Poor
Oral Antihistamines	1	8	1		10	PA
No Antihistamines	4	6			10	PA
IM Antinauseants			1	9	10	Poor
Oral Antinauseants	5		1	4	10	PA
No Antinauseants	6	4			10	High
IM Narcotic Analgesics	1		1	8	10	Poor
Oral Narcotic Analgesics	2		3	5	10	PNA
No Narcotic Analgesics	8	2			10	High
IM Non-narcotic Analgesics	1		1	8	10	Poor
Oral Non-narcotic Analgesics	5	5			10	High
No Non-narcotic Analgesics	4	5			9 ^e	PA
Oral Decongestants	2	7	1		10	PA
No Decongestants	3	4	2	1	10	PA
Bronchial Dilators		2	2	6	10	Poor
No Bronchial Dilators	7	2		1	10	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

211-
 Table B.4, cont.
OTITIS MEDIA 5-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray						
No Chest X-ray	8	1	4	5	10	PNA High
Sinus X-ray						
No Sinus X-ray	6	3	6	1	10	PNA High
WBC/CBC						
No WBC/CBC	4	6	4		10	PA PA
Heterophile						
No Heterophile	4	1	7	2	10	PNA PA
Urinalysis						
No Urinalysis	4	2	6	2	10	PNA PA
Strep/Throat Culture						
No Strep/Throat Culture	4	7	2	1	10	PA PA
Other Culture						
No Other Culture	2	7	1		10	PA PA
Panel/Profile						
No Panel/Profile	5	1	5	4	10	PNA High
Followup Visits						
No Followup Visits	9	1		3	10	High Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.5
PHARYNGITIS 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None	3	4	1	2	10	PA
IM SA Penicillin			3	7	10	Poor
IM LA Penicillin	4	5	1		10	PA
IM Ampicillin		3	3	4	10	PNA
IM Lincomycin			1	9	10	Poor
IM Tetracycline		1		9	10	Poor
Oral Penicillin	4	5	1		10	PA
Oral Ampicillin		4	6		10	PNA
Oral Lincomycin		1	1	8	10	Poor
Oral Tetracycline		2		8	10	Poor
Oral Erythromycin	2	7	1		10	PA
Oral Cephalosporin		3	2	5	10	PNA
Oral Sulfa			3	7	10	Poor
IM SA Pen + Oral Pen	2	6	1	1	10	PA
IM SA Pen + Oral Amp		4	5	1	10	PNA
IM SA Pen + Oral Tetra	1		9		10	Poor
IM SA Pen + Oral Erythrom		5	2	3	10	PA
IM SA Pen + Oral Lincom			1	9	10	Poor
IM LA Pen + Oral Pen	1	3	3	3	10	PNA
IM LA Pen + Oral Amp		2	5	3	10	PNA
IM LA Pen + Oral Erythrom		2	5	3	10	PNA
IM LA Pen + Oral Lincom			2	8	10	Poor
IM LA Pen + Oral Tetra	1	1	8		10	Poor
IM LA Pen + Oral Sulfa	1	3	6		10	Poor
IM Amp + Oral Pen		3	4	3	10	PNA
IM Amp + Oral Amp		2	5	3	10	PNA
IM Lincom + Oral Pen			2	8	10	Poor
IM Lincom + Oral Amp			2	8	10	Poor
IM Lincom + Oral Tetra		1	9		10	Poor
IM Lincom + Oral Erythrom		1	8	9 ^e	10	Poor
IM Lincom + Oral Lincom			1	9	10	Poor
IM Tetra + Oral Tetra		1	9	10		Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

213
Table B.5, cont.
PHARYNGITIS 0-7 years

<u>TYPE OF DRUG OR SERVICE</u>	<u>NUMBER OF JUDGES RATING EACH SERVICE AS:</u>					<u>MEDIAN CATEGORY^c</u>
	<u>High</u>	<u>PA^a</u>	<u>PNA^a</u>	<u>Poor</u>	<u>Total^b</u>	
Antibiotics						
IM Antitussives				10	10	Poor
Oral Antitussives	7	2	1	10	PA	
No Antitussives	6	4		10	High	
IM Antihistamines				10	10	Poor
Oral Antihistamines	4	4	2	10	PNA	
No Antihistamines	7	3		10	High	
IM Antinauseants			1	9	10	Poor
Oral Antinauseants	2	5	3	10	PNA	
No Antinauseants	6	3	1	10	High	
IM Narcotic Analgesics				10	10	Poor
Oral Narcotic Analgesics			2	8	10	Poor
No Narcotic Analgesics	9			1	10	High
IM Non-narcotic Analgesics			1	9	10	Poor
Oral Non-narcotic Analgesics	2	6	2	2	10	PA
No Non-narcotic Analgesics	4	5	1		10	PA
Oral Decongestants		6	1	1	10	PA
No Decongestants	5	4	1		10	High
Bronchial Dilators		2	2	6	10	Poor
No Bronchial Dilators	6	3		1	10	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.5, cont.

PHARYNGITIS 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray		2	5	3	10	PNA
No Chest X-ray	7	3			10	High
Sinus X-ray		2	3	5	10	PNA
No Sinus X-ray	8	2			10	High
WBC/CBC		6	4		10	PA
No WBC/CBC	5	5			10	High
Heterophile	1	6	2	1	10	PA
No Heterophile	6	4			10	High
Urinalysis		4	3	3	10	PNA
No Urinalysis	7	2	1		10	High
Strep/Throat Culture	9	1			10	High
No Strep/Throat Culture	1	4	3	2	10	PA
Other Culture		3	4	3	10	PNA
No Other Culture	7	3			10	High
Panel/Profile		1	4	5	10	PNA
No Panel/Profile	6	4			10	High
Followup Visits	1	7	2		10	PA
No Followup Visits	2	8			10	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.6

PHARYNGITIS 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None	3	5	1	2	11	PA
IM SA Penicillin		1	3	7	11	Poor
IM LA Penicillin	4	6	1		11	PA
IM Ampicillin		3	4	4	11	PNA
IM Lincomycin			2	9	11	Poor
IM Tetracycline		2	1	8	11	Poor
Oral Penicillin	4	6	1		11	PA
Oral Ampicillin		3	7	1	11	PNA
Oral Lincomycin		1	2	8	11	Poor
Oral Tetracycline		3	2	6	11	Poor
Oral Erythromycin	3	7	1		11	PA
Oral Cephalosporin		4	2	5	11	PNA
Oral Sulfa		1	3	7	11	Poor
IM SA Pen + Oral Pen	2	7	1	1	11	PA
IM SA Pen + Oral Amp		5	3	3	11	PNA
IM SA Pen + Oral Tetra		2	1	8	11	Poor
IM SA Pen + Oral Erythrom		6	2	3	11	PA
IM SA Pen + Oral Lincom			2	9	11	Poor
IM LA Pen + Oral Pen	1	4	3	3	11	PNA
IM LA Pen + Oral Amp		3	5	3	11	PNA
IM LA Pen + Oral Erythrom		3	5	3	11	PNA
IM LA Pen + Oral Lincom			3	8	11	Poor
IM LA Pen + Oral Tetra	2	2	7		11	Poor
IM LA Pen + Oral Sulfa	1	4	6		11	Poor
IM Amp + Oral Pen		3	4	4	11	PNA
IM Amp + Oral Amp	2	5	4		11	PNA
IM Lincom + Oral Pen		3	8		11	Poor
IM Lincom + Oral Amp		3	8		11	Poor
IM Lincom + Oral Tetra		2	9		11	Poor
IM Lincom + Oral Erythrom		2	8		10 ^e	Poor
IM Lincom + Oral Lincom		1	10		11	Poor
IM Tetra + Oral Tetra	1		10	11		Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.6, cont.
PHARYNGITIS 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives	1			10	11	Poor
Oral Antitussives	7	2		2	11	PA
No Antitussives	7	3	1		11	High
IM Antihistamines		1		10	11	Poor
Oral Antihistamines		4	4	3	11	PNA
No Antihistamines	8	2	1		11	High
IM Antinauseants		2		9	11	Poor
Oral Antinauseants		2	6	3	11	PNA
No Antinauseants	7	3	1		11	High
IM Narcotic Analgesics				11	11	Poor
Oral Narcotic Analgesics		1	2	8	11	Poor
No Narcotic Analgesics	9	1	1		11	High
IM Non-narcotic Analgesics		2		9	11	Poor
Oral Non-narcotic Analgesics	2	7	1	1	11	PA
No Non-narcotic Analgesics	5	5	1		11	PA
Oral Decongestants	1	6	2	2	11	PA
No Decongestants	6	4	1		11	High
Bronchial Dilators		2	2	7	11	Poor
No Bronchial Dilators	7	2	2		11	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.6, cont.
PHARYNGITIS 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray		1	6	4	11	PNA
No Chest X-ray	9	2			11	High
Sinus X-ray		2	4	5	11	PNA
No Sinus X-ray	9	2			11	High
WBC/CBC		5	6		11	PNA
No WBC/CBC	6	4	1		11	High
Heterophile		2	6	3	11	PNA
No Heterophile	6	5			11	High
Urinalysis		2	6	3	11	PNA
No Urinalysis	8	2	1		11	High
Strep/Throat Culture		7	4		11	High
No Strep/Throat Culture	1	5	4	1	11	PA
Other Culture		3	5	3	11	PNA
No Other Culture	8	3			11	High
Panel/Profile		1	5	5	11	PNA
No Panel/Profile	7	4			11	High
Followup Visits	1	7	3		11	PA
No Followup Visits	3	8			11	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

218
Table B.7

BRONCHITIS 0-7 years

<u>TYPE OF DRUG OR SERVICE</u>	<u>NUMBER OF JUDGES RATING EACH SERVICE AS:</u>					<u>MEDIAN CATEGORY^c</u>
	<u>High</u>	<u>PA^a</u>	<u>PNA^a</u>	<u>Poor</u>	<u>Total^b</u>	
Antibiotics						
None	4	6			10	PA
IM SA Penicillin		6	4	10		PNA
IM LA Penicillin	2	5	3	10		PNA
IM Ampicillin	1	2	3	4	10	PNA
IM Lincomycin		1	9	10		Poor
IM Tetracycline		1	9	10		Poor
Oral Penicillin	1	4	3	2	10	PA
Oral Ampicillin	1	7	1	1	10	PA
Oral Lincomycin		1	9	10		Poor
Oral Tetracycline		1	8	9 ^e		Poor
Oral Erythromycin	1	3	2	3	9 ^e	PNA
Oral Cephalosporin		1	3	6	10	Poor
Oral Sulfa		3	7	10		Poor
IM SA Pen + Oral Pen	1	3	3	3	10	PNA
IM SA Pen + Oral Amp	1	3	2	4	10	PNA
IM SA Pen + Oral Tetra		1	9	10		Poor
IM SA Pen + Oral Erythrom	1	1	4	4	10	PNA
IM SA Pen + Oral Lincom		1	9	10		Poor
IM LA Pen + Oral Pen		3	3	4	10	PNA
IM LA Pen + Oral Amp	4	2	4	4	10	PNA
IM LA Pen + Oral Erythrom	3	3	4	4	10	PNA
IM LA Pen + Oral Lincom		1	9	10		Poor
IM LA Pen + Oral Tetra	1		9	10		Poor
IM LA Pen + Oral Sulfa	1	3	6	10		Poor
IM Amp + Oral Pen		2	3	5	10	PNA
IM Amp + Oral Amp	1	4	2	3	10	PA
IM Lincom + Oral Pen		1	9	10		Poor
IM Lincom + Oral Amp		1	9	10		Poor
IM Lincom + Oral Tetra		1	9	10		Poor
IM Lincom + Oral Erythrom		1	9	10		Poor
IM Lincom + Oral Lincom		1	9	10		Poor
IM Tetra + Oral Tetra		1	9	10		Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.7, cont.
BRONCHITIS 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives		1		9	10	Poor
Oral Antitussives	2	6	1	1	10	PA
No Antitussives	4	5	1		10	PA
IM Antihistamines		1	1	8	10	Poor
Oral Antihistamines	1	3	3	3	10	PNA
No Antihistamines	4	5		1	10	PA
IM Antinauseants		1		9	10	Poor
Oral Antinauseants		3	2	5	10	PNA
No Antinauseants	4	5		1	10	PA
IM Narcotic Analgesics				10	10	Poor
Oral Narcotic Analgesics		1		9	10	Poor
No Narcotic Analgesics	6	3		1	10	High
IM Non-narcotic Analgesics		2	1	7	10	Poor
Oral Non-narcotic Analgesics		8		2	10	PA
No Non-narcotic Analgesics	3	7			10	PA
Oral Decongestants		7	1	2	10	PA
No Decongestants	4	6			10	PA
Bronchial Dilators	1	8	1		10	PA
No Bronchial Dilators	2	7	1		10	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

220
Table B.7, cont.
BRONCHITIS 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray	4	5	1		10	PA
No Chest X-ray	2	4	4		10	PA
Sinus X-ray		3	6	1	10	PNA
No Sinus X-ray	4	6			10	PA
WBC/CBC	1	9			10	PA
No WBC/CBC	1	8	1		10	PA
Heterophile			8	2	10	PNA
No Heterophile	4	6			10	PA
Urinalysis		3	3	4	10	PNA
No Urinalysis	6	4			10	High
Strep/Throat Culture	1	5	3	1	10	PA
No Strep/Throat Culture	5	4		1	10	High
Other Culture	3	4	3		10	PA
No Other Culture	2	7		1	10	PA
Panel/Profile		1	5	4	10	PNA
No Panel/Profile	2	7			9 ^e	PA
Followup Visits	5	4	1		10	High
No Followup Visits	1	3	3	3	10	PNA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

221
 Table B.8
BRONCHITIS 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None	4	6	1		11	PA
IM SA Penicillin		1	6	4	11	PNA
IM LA Penicillin		3	5	3	11	PNA
IM Ampicillin	1	3	3	4	11	PNA
IM Lincomycin			2	9	11	Poor
IM Tetracycline		2	2	7	11	Poor
Oral Penicillin	1	5	4	1	11	PA
Oral Ampicillin	2	7	1	1	11	PA
Oral Lincomycin			2	9	11	Poor
Oral Tetracycline		7	1	3	11	PA
Oral Erythromycin	2	4	1	3	10 ^e	PA
Oral Cephalosporin		4	2	5	11	PNA
Oral Sulfa		1	4	6	11	Poor
IM SA Pen + Oral Pen	1	4	3	3	11	PNA
IM SA Pen + Oral Amp	1	4	2	4	11	PNA
IM SA Pen + Oral Tetra		3	1	7	11	Poor
IM SA Pen + Oral Erythrom	2	3	1	5	11	PNA
IM SA Pen + Oral Lincom			2	9	11	Poor
IM LA Pen + Oral Pen		5	2	4	11	PNA
IM LA Pen + Oral Amp	1	4	2	4	11	PNA
IM LA Pen + Oral Erythrom	1	4	2	4	11	PNA
IM LA Pen + Oral Lincom			2	9	11	Poor
IM LA Pen + Oral Tetra		3	1	7	11	Poor
IM LA Pen + Oral Sulfa		1	4	6	11	Poor
IM Amp + Oral Pen		4	1	6	11	Poor
IM Amp + Oral Amp	2	4	1	4	11	PA
IM Lincom + Oral Pen			2	9	11	Poor
IM Lincom + Oral Amp		2	9	11		Poor
IM Lincom + Oral Tetra		2	9	11		Poor
IM Lincom + Oral Erythrom		2	9	11		Poor
IM Lincom + Oral Lincom		1	10	11		Poor
IM Tetra + Oral Tetra	1	2	8	11		Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.8, cont.
BRONCHITIS 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives	2	1	8	11		Poor
Oral Antitussives	2	8	1	11		PA
No Antitussives	5	5	1	11		PA
IM Antihistamines	1	2	8	11		Poor
Oral Antihistamines	1	3	3	11		PNA
No Antihistamines	5	4	2	11		PA
IM Antinauseants	1	1	9	11		Poor
Oral Antinauseants	3	4	4	11		PNA
No Antinauseants	4	6	1	11		PA
IM Narcotic Analgesics		1	10	11		Poor
Oral Narcotic Analgesics		1	2	11		Poor
No Narcotic Analgesics	7	3	1	11		High
IM Non-narcotic Analgesics	2	2	7	11		Poor
Oral Non-narcotic Analgesics	9		2	11		PA
No Non-narcotic Analgesics	4	7		11		PA
Oral Decongestants	1	7	1	11		PA
No Decongestants	5	6		11		PA
Bronchial Dilators	1	8	2	11		PA
No Bronchial Dilators	3	7	1	11		PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.8, cont.

BRONCHITIS 8+ years

<u>TYPE OF DRUG OR SERVICE</u>	<u>NUMBER OF JUDGES RATING EACH SERVICE AS:</u>					<u>MEDIAN CATEGORY^c</u>
	<u>High</u>	<u>PA^a</u>	<u>PNA^a</u>	<u>Poor</u>	<u>Total^b</u>	
Antibiotics						
Chest X-ray		2	7	2	11	PNA
No Chest X-ray	3	5	2	1	11	PA
Sinus X-ray		4	6	1	11	PNA
No Sinus X-ray	5	6			11	PA
WBC/CBC		10	1		11	PA
No WBC/CBC	2	9			11	PA
Heterophile			9	2	11	PNA
No Heterophile	5	6			11	PA
Urinalysis		2	5	4	11	PNA
No Urinalysis	6	4	1		11	High
Strep/Throat Culture	2	4	4	1	11	PA
No Strep/Throat Culture	6	4		1	11	High
Other Culture	3	4	4		11	PA
No Other Culture	3	7		1	11	PA
Panel/Profile		1	6	4	11	PNA
No Panel/Profile	6	4			10 ^e	High
Followup Visits	5	5	1		11	PA
No Followup Visits	1	5	2	3	11	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

224
 Table B.9
INFLUENZA 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None	9	1			10	High
IM SA Penicillin		3	7	10		Poor
IM LA Penicillin		3	7	10		Poor
IM Ampicillin		4	6	10		Poor
IM Lincomycin		1	9	10		Poor
IM Tetracycline		1	9	10		Poor
Oral Penicillin		5	5	10		PNA
Oral Ampicillin	1	4	5	10		Poor
Oral Lincomycin		1	9	10		Poor
Oral Tetracycline		1	8	9 ^e		Poor
Oral Erythromycin		3	6	9 ^e		Poor
Oral Cephalosporin		2	8	10		Poor
Oral Sulfa		3	7	10		Poor
IM SA Pen + Oral Pen		4	6	10		Poor
IM SA Pen + Oral Amp		3	7	10		Poor
IM SA Pen + Oral Tetra		1	9	10		Poor
IM SA Pen + Oral Erythrom		3	7	10		Poor
IM SA Pen + Oral Lincom		1	9	10		Poor
IM LA Pen + Oral Pen		3	7	10		Poor
IM LA Pen + Oral Amp		3	7	10		Poor
IM LA Pen + Oral Erythrom		3	7	10		Poor
IM LA Pen + Oral Lincom		1	9	10		Poor
IM LA Pen + Oral Tetra		1	9	10		Poor
IM LA Pen + Oral Sulfa		3	7	10		Poor
IM Amp + Oral Pen		3	7	10		Poor
IM Amp + Oral Amp		4	6	10		Poor
IM Lincom + Oral Pen		2	8	10		Poor
IM Lincom + Oral Amp		2	8	10		Poor
IM Lincom + Oral Tetra		1	9	10		Poor
IM Lincom + Oral Erythrom		2	8	10		Poor
IM Lincom + Oral Lincom		1	9	10		Poor
IM Tetra + Oral Tetra		2	8	10		Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.9, cont.
INFLUENZA 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives		1	1	8	10	Poor
Oral Antitussives	1	8	1		10	PA
No Antitussives	3	7			10	PA
IM Antihistamines		1		9	10	Poor
Oral Antihistamines		6	1	3	10	PA
No Antihistamines	5	4	1		10	High
IM Antinauseants		1	1	8	10	Poor
Oral Antinauseants		6	1	3	10	PA
No Antinauseants	3	5		2	10	PA
IM Narcotic Analgesics				10	10	Poor
Oral Narcotic Analgesics		2		8	10	Poor
No Narcotic Analgesics	8	1	1		10	High
IM Non-narcotic Analgesics		2		8	10	Poor
Oral Non-narcotic Analgesics	3	5		2	10	PA
No Non-narcotic Analgesics	2	8			10	PA
Oral Decongestants		7	3		10	PA
No Decongestants	3	7			10	PA
Bronchial Dilators		3	5	2	10	PNA
No Bronchial Dilators	4	6			10	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.9, cont.
INFLUENZA 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray		7	3		10	PA
No Chest X-ray	3	7			10	PA
Sinus X-ray		2	6	2	10	PNA
No Sinus X-ray	4	6			10	PA
WBC/CBC		7	3		10	PA
No WBC/CBC	2	8			10	PA
Heterophile		1	6	3	10	PNA
No Heterophile	3	7			10	PA
Urinalysis		4	4	2	10	PNA
No Urinalysis	2	7	1		10	PA
Strep/Throat Culture		7	2	1	10	PA
No Strep/Throat Culture	2	7	1		10	PA
Other Culture		3	6	1	10	PNA
No Other Culture	2	7	1		10	PA
Panel/Profile		1	6	3	10	PNA
No Panel/Profile	3	5	1		10	PA
Followup Visits	2	7	1		10	PA
No Followup Visits	2	7	1		10	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.10

INFLUENZA 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None	10	1			11	High
IM SA Penicillin	1	3	7	11		Poor
IM LA Penicillin	1	3	7	11		Poor
IM Ampicillin	1	4	6	11		Poor
IM Lincomycin		2	9	11		Poor
IM Tetracycline	1	3	7	11		Poor
Oral Penicillin	1	5	5	11		PNA
Oral Ampicillin	2	4	5	11		PNA
Oral Lincomycin		2	9	11		Poor
Oral Tetracycline	2	3	6	11		Poor
Oral Erythromycin	1	3	6	10 ^e		Poor
Oral Cephalosporin	1	2	8	11		Poor
Oral Sulfa	1	3	7	11		Poor
IM SA Pen + Oral Pen	1	4	6	11		Poor
IM SA Pen + Oral Amp	1	3	7	11		Poor
IM SA Pen + Oral Tetra	1	3	7	11		Poor
IM SA Pen + Oral Erythrom	1	3	7	11		Poor
IM SA Pen + Oral Lincom		2	9	11		Poor
IM LA Pen + Oral Pen	1	3	7	11		Poor
IM LA Pen + Oral Amp	1	3	7	11		Poor
IM LA Pen + Oral Erythrom	1	3	7	11		Poor
IM LA Pen + Oral Lincom		2	9	11		Poor
IM LA Pen + Oral Tetra	1	3	7	11		Poor
IM LA Pen + Oral Sulfa	1	3	7	11		Poor
IM Amp + Oral Pen	1	3	7	11		Poor
IM Amp + Oral Amp	1	4	6	11		Poor
IM Lincom + Oral Pen		3	8	11		Poor
IM Lincom + Oral Amp		3	8	11		Poor
IM Lincom + Oral Tetra		3	8	11		Poor
IM Lincom + Oral Erythrom		3	8	11		Poor
IM Lincom + Oral Lincom		1	10	11		Poor
IM Tetra + Oral Tetra		4	7	11		Poor

^a PA = Probably Acceptable; PNA = Probably Not Acceptable

^b The total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^c In the event of "ties," the higher rating was used.

^d One judge did not respond. (See text for explanation.)

Table B.10, cont.

INFLUENZA 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives	2	1	8	11		Poor
Oral Antitussives	1	9	1	11		PA
No Antitussives	4	7		11		PA
IM Antihistamines	1	1	9	11		Poor
Oral Antihistamines	6	1	4	11		PA
No Antihistamines	6	3	2	11		High
IM Antinauseants	1	2	8	11		Poor
Oral Antinauseants	5	2	4	11		PNA
No Antinauseants	4	5	2	11		PA
IM Narcotic Analgesics		1	10	11		Poor
Oral Narcotic Analgesics	2	1	8	11		Poor
No Narcotic Analgesics	8	2	1	11		High
IM Non-narcotic Analgesics	2	1	8	11		Poor
Oral Non-narcotic Analgesics	4	5	2	11		PA
No Non-narcotic Analgesics	3	8		11		PA
Oral Decongestants	1	9	1	11		PA
No Decongestants	4	7		11		PA
Bronchial Dilators		4	5	2	11	PNA
No Bronchial Dilators	5	6		11		PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.10, cont.
INFLUENZA 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray		6	5		11	PA
No Chest X-ray	4	7			11	PA
Sinus X-ray		2	7	2	11	PNA
No Sinus X-ray	5	6			11	PA
WBC/CBC		7	4		11	PA
No WBC/CBC	3	8			11	PA
Heterophile		2	6	3	11	PNA
No Heterophile	4	7			11	PA
Urinalysis		3	4	4	11	PNA
No Urinalysis	4	6	1		11	PA
Strep/Throat Culture		7	3		10 ^e	PA
No Strep/Throat Culture	4	6			10 ^e	PA
Other Culture		2	7	2	11	PNA
No Other Culture	5	6			11	PA
Panel/Profile			7	4	11	PNA
No Panel/Profile	6	4			10 ^e	High
Followup Visits	2	7	1	1	11	PA
No Followup Visits	3	7	1		11	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.11
ACUTE URI 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None	10				10	High
IM SA Penicillin		2	8	10		Poor
IM LA Penicillin		2	8	10		Poor
IM Ampicillin		2	8	10		Poor
IM Lincomycin		1	9	10		Poor
IM Tetracycline		1	9	10		Poor
Oral Penicillin	1	3	6	10		Poor
Oral Ampicillin	1	3	6	10		Poor
Oral Lincomycin		1	9	10		Poor
Oral Tetracycline		1	9	10		Poor
Oral Erythromycin	1	3	6	10		Poor
Oral Cephalosporin		2	8	10		Poor
Oral Sulfa		3	7	10		Poor
IM SA Pen + Oral Pen		2	8	10		Poor
IM SA Pen + Oral Amp		2	8	10		Poor
IM SA Pen + Oral Tetra			10	10		Poor
IM SA Pen + Oral Erythrom		2	8	10		Poor
IM SA Pen + Oral Lincom		1	9	10		Poor
IM LA Pen + Oral Pen		2	8	10		Poor
IM LA Pen + Oral Amp		2	8	10		Poor
IM LA Pen + Oral Erythrom		2	8	10		Poor
IM LA Pen + Oral Lincom		1	9	10		Poor
IM LA Pen + Oral Tetra			10	10		Poor
IM LA Pen + Oral Sulfa		2	8	10		Poor
IM Amp + Oral Pen		2	8	10		Poor
IM Amp + Oral Amp		2	8	10		Poor
IM Lincom + Oral Pen		2	8	10		Poor
IM Lincom + Oral Amp		2	8	10		Poor
IM Lincom + Oral Tetra			10	10		Poor
IM Lincom + Oral Erythrom		2	8	10		Poor
IM Lincom + Oral Lincom		1	9	10		Poor
IM Tetra + Oral Tetra			10	10		Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.11, cont.
ACUTE URI 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
IM Antitussives			2	8	10	Poor
Oral Antitussives	2	4	4		10	PA
No Antitussives	6	4			10	High
IM Antihistamines			2	8	10	Poor
Oral Antihistamines	1	7	2		10	PA
No Antihistamines	5	5			10	High
IM Antinauseants			2	8	10	Poor
Oral Antinauseants		3	2	5	10	PNA
No Antinauseants	4	4		2	10	PA
IM Narcotic Analgesics			1	9	10	Poor
Oral Narcotic Analgesics		1	1	8	10	Poor
No Narcotic Analgesics	7	2		1	10	High
IM Non-narcotic Analgesics			1	8	9 ^e	Poor
Oral Non-narcotic Analgesics	1	6	1	2	10	PA
No Non-narcotic Analgesics	3	7			10	PA
Oral Decongestants	1	8	1		10	PA
No Decongestants	3	7			10	PA
Bronchial Dilators		2	5	3	10	PNA
No Bronchial Dilators	6	4			10	High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.11, cont.
ACUTE URI 0-7 years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
Chest X-ray		1	8	1	10	PNA
No Chest X-ray	6	4			10	High
Sinus X-ray		2	5	3	10	PNA
No Sinus X-ray	5	5			10	High
WBC/CBC		4	6		10	PNA
No WBC/CBC	4	6			10	PA
Heterophile		2	4	4	10	PNA
No Heterophile	6	3		1	10	High
Urinalysis		2	4	4	10	PNA
No Urinalysis	6	4			10	High
Strep/Throat Culture		7	3		10	PA
No Strep/Throat Culture	6	4			10	High
Other Culture		4	5	1	10	PNA
No Other Culture	4	6			10	PA
Panel/Profile		1	4	5	10	PNA
No Panel/Profile	6	3			9 ^e	High
Followup Visits	1	3	6		10	PNA
No Followup Visits	4	5	1		10	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.12

ACUTE URI 8+ years

TYPE OF DRUG OR SERVICE	NUMBER OF JUDGES RATING EACH SERVICE AS:					MEDIAN CATEGORY ^c
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
None	10	1			11	High
IM SA Penicillin	1	2	8	11		Poor
IM LA Penicillin	1	2	8	11		Poor
IM Ampicillin	1	2	8	11		Poor
IM Lincomycin		2	9	11		Poor
IM Tetracycline	1	2	8	11		Poor
Oral Penicillin	1	3	7	11		Poor
Oral Ampicillin	1	3	7	11		Poor
Oral Lincomycin		2	9	11		Poor
Oral Tetracycline	1	3	7	11		Poor
Oral Erythromycin	1	3	7	11		Poor
Oral Cephalosporin	1	2	8	11		Poor
Oral Sulfa	1	3	7	11		Poor
IM SA Pen + Oral Pen	1	2	8	11		Poor
IM SA Pen + Oral Amp	1	2	8	11		Poor
IM SA Pen + Oral Tetra	1	2	8	11		Poor
IM SA Pen + Oral Erythrom	1	2	8	11		Poor
IM SA Pen + Oral Lincom		2	9	11		Poor
IM LA Pen + Oral Pen	1	2	8	11		Poor
IM LA Pen + Oral Amp	1	2	8	11		Poor
IM LA Pen + Oral Erythrom	1	2	8	11		Poor
IM LA Pen + Oral Lincom		2	9	11		Poor
IM LA Pen + Oral Tetra	1	2	8	11		Poor
IM LA Pen + Oral Sulfa	1	2	8	11		Poor
IM Amp + Oral Pen	1	2	8	11		Poor
IM Amp + Oral Amp	1	2	8	11		Poor
IM Lincom + Oral Pen	3	8	11			Poor
IM Lincom + Oral Amp	3	8	11			Poor
IM Lincom + Oral Tetra	3	8	11			Poor
IM Lincom + Oral Erythrom	3	8	11			Poor
IM Lincom + Oral Lincom	1	10	11			Poor
IM Tetra + Oral Tetra	3	8	11			Poor

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

234
Table B.12, cont.

ACUTE URI 8+ years

<u>TYPE OF DRUG OR SERVICE</u>	<u>NUMBER OF JUDGES RATING EACH SERVICE AS:</u>					<u>MEDIAN CATEGORY^c</u>
	High	PA ^a	PNA ^a	Poor	Total ^b	
Antibiotics						
IM Antitussives	1	2	8	11		Poor
Oral Antitussives	2	5	4	11		PA
No Antitussives	7	4		11		High
IM Antihistamines	1	2	8	11		Poor
Oral Antihistamines	2	7	2	11		PA
No Antihistamines	6	5		11		High
IM Antinauseants	1	2	8	11		Poor
Oral Antinauseants	4	2	5	11		PNA
No Antinauseants	5	4	2	11		PA
IM Narcotic Analgesics		2	9	11		Poor
Oral Narcotic Analgesics	1	2	8	11		Poor
No Narcotic Analgesics	8	2	1	11		High
IM Non-narcotic Analgesics		2	8	10 ^e		Poor
Oral Non-narcotic Analgesics	1	7	2	11		PA
No Non-narcotic Analgesics	4	7		11		PA
Oral Decongestants	3	7	1	11		PA
No Decongestants	4	7		11		PA
Bronchial Dilators	1	7	3	11		PNA
No Bronchial Dilators	7	4		11		High

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

Table B.12
ACUTE URI 8+ years

<u>TYPE OF DRUG OR SERVICE</u>	<u>NUMBER OF JUDGES RATING EACH SERVICE AS:</u>					<u>MEDIAN CATEGORY^c</u>
	<u>High</u>	<u>PA^a</u>	<u>PNA^a</u>	<u>Poor</u>	<u>Total^b</u>	
Chest X-ray			10	1	11	PNA
No Chest X-ray	7	4			11	High
Sinus X-ray		3	5	3	11	PNA
No Sinus X-ray	6	5			11	High
WBC/CBC		3	6	1	10 ^e	PNA
No WBC/CBC	5	5			10 ^e	High
Heterophile		3	6	2	11	PNA
No Heterophile	6	5			11	High
Urinalysis		1	6	4	11	PNA
No Urinalysis	7	4			11	High
Strep/Throat Culture		6	5		11	PA
No Strep/Throat Culture	7	4			11	High
Other Culture		3	6	2	11	PNA
No Other Culture	5	6			11	PA
Panel/Profile		1	5	5	11	PNA
No Panel/Profile	7	3			10 ^e	High
Followup Visits	1	4	6		11	PNA
No Followup Visits	5	5	1		11	PA

^aPA = Probably Acceptable; PNA = Probably Not Acceptable

^bThe total number of judges for the pediatric age group was 10; the total for the 8 years and older group was 11.

^cIn the event of "ties," the higher rating was used.

^dOne judge did not respond. (See text for explanation.)

APPENDIX C
EPISODE RULES

The final set of rules to create episodes is described below. The steps to reach the final workfile are noted first, followed by the general and diagnosis-specific rules and a note about antibiotic variables.

PRODUCE FINAL WORKFILE

These analyses are based on the same special ambulatory care data tape that was prepared by the EMCRO's fiscal intermediary and used for the previous episode analyses (see Chapter I). The first steps in preparing a working tape for these analyses were the following: First, data were extracted from the original Dikewood tape for only Medicaid eligibles who had been enrolled in the Aid to Families with Dependent Children (AFDC) program continuously for four years. This formed the cohort used in all previous studies.

Second, all data were extracted for only those persons who had had at least one visit for one of 11 infectious disease diagnoses (the six study diagnoses and five additional closely related diagnoses that were shown in Table IV.1). This was done for each period separately.

Third, all data falling within specific ranges of procedure, visit, or laboratory test codes or codes for injectable or oral medications were extracted. These were restricted to services believed to be relevant to this study; a need for economy in computing costs and for

simplifying the quality-of-care assessment dictated eliminating from the data base services that normally would appear very infrequently for these conditions, rather than resorting to ever-more-complex rules or quality-of-care criteria. The specific types of services retained in the data base included the following:

PROCEDURES: Chest x-rays; sinus x-rays; complete and white blood counts; other blood tests (hematocrit, hemoglobin, hemogram, heterophile, erythrocyte count, and peripheral smear); urinalyses and urine cultures; throat and strep cultures; cultures from other sites (sputum, nasal passages, ear, etc.); and other tests (largely panel and profile tests that might include one or more of the above tests, and sensitivity studies).

PHYSICIAN VISITS: All procedure codes for physician visits were included. This specifically means all physician office visits for new and established patients, home visits for new and established patients, and emergency room and outpatient department visits to inpatient facilities. The range includes minimal, brief, limited, intermediate, extended, and comprehensive, as those levels of care are typically understood in California Relative Values Study terminology.

PRESCRIPTION DRUGS: Antibiotic prescriptions included all types of penicillin and ampicillin, lincomycin and clindamycin, tetracycline, erythromycin, cephalosporin, sulfa drugs, and various other antibiotics including gram-negative drugs, and antifungals. All other oral medications included in these analyses can be considered basically symptomatic--i.e., drugs intended to relieve symptoms and discomfort, not to cure disease. They include antitussives and expectorants,

antihistamines, antinauseants and antiemetics, narcotic analgesics, non-narcotic analgesics, various other nasal preparations, and bronchial dilators. The distinction between antihistamines and antitussives was sometimes difficult to make because of the existence of combination preparations that are often widely used; in these circumstances, the therapeutic distinctions used by Dikewood (as developed by Dr. Atkinson from the Department of Health and Social Services in New Mexico) were maintained. Narcotic and non-narcotic analgesics correspond roughly to the distinction maintained in the AMA Drug Evaluations of strong and mild analgesics, respectively. Nasal preparations can include nasal antibiotics, various cortisone preparations, and a variety of other drugs of that nature. Decongestants fall basically into the category of antitussives and expectorants.

INJECTABLE DRUGS: Antibiotics include the specific categories of short-acting penicillin (e.g., procaine penicillin G); long-acting penicillin (benzathine penicillin G, commonly referred to as Bicillin); ampicillin; lincomycin; and tetracycline. All other antibiotics and combinations were grouped together and include small numbers of IM drugs such as erythromycin, aminoglycosides, and cephalosporins, as well as certain antibiotics combinations and antibiotic/nonantibiotic combinations. Other injectable drugs included for these analyses are antihistamines, antinauseants and antiemetics, antitussives, narcotic analgesics, and non-narcotic analgesics. As elsewhere, injectable here refers to intramuscular, or IM, drugs.

The steps of extracting data onto a working file that included only information with specific ranges of codes for procedures and medications

relevant to these studies included a substantial amount of collapsing and/or disaggregating the original codes as they appeared on the Dikewood tape. By and large, this recoding involved aggregating numerous procedure codes for, e.g., blood tests into a single category and aggregating codes for IM drugs into fewer but nonetheless homogeneous groupings. With regard to the codes for prescription drugs, in some cases (mainly antibiotics) the therapeutic codes used by Dikewood were too inclusive, and disaggregated categories were developed; in other instances, oral drugs were grouped into larger classifications. Again, these types of recodes had been done for the earlier studies and only a few new categories needed to be recreated from the existing ones.

Fourth, all claims that contained two diagnoses were recorded to a single diagnosis, namely the one that was considered to be more bacterial according to the original hierarchy. With one exception, this also had been done for the earlier analyses. The exception was that all claims involving any diagnosis of infectious mononucleosis as one of a pair were left as that diagnosis, so that episodes ostensibly of a respiratory infection that in fact were likely to be the more serious illness (infectious mononucleosis) could be dropped from the analyses.

Fifth, some of the above medications and services had been recoded into more (or less) aggregated groupings for the earlier studies. For the present study, some of these groupings were changed, to make them more homogenous and in conformance with the specifications of the quality-of-care criteria.

CREATE EPISODESGeneral Rules

For every study diagnosis except otitis media, episodes were created on a single pass through the tape and recorded in the order in which they occurred for each patient (recalling that the tape was ordered by patient ID). Otitis media episodes were created on a separate run through the tape, and the rules governing otitis media are dealt with separately below. Study diagnosis in this section thus refers to strep throat, pharyngitis/tonsillitis, acute bronchitis, influenza, and acute URI.

The first visit with one of the study diagnoses encountered for each patient began an episode, and the Day on which it occurred was considered "Day 1." Generally, all visits, laboratory tests, procedures, and medications in both oral and injectable (IM) form that occurred on Day 1 through Day 15 (i.e., in a two-week period) and that were given for the diagnosis of the episode or for a closely related diagnosis were included in the episode. Closely related diagnoses are those denoted CR on Table IV.1 and the "other five" study diagnoses (with respect to any given study diagnosis).

Episodes of all diagnoses were constrained not to begin within two weeks of the beginning of either Period I or II and not to begin with two weeks of the end of either Period I or II. (This helped eliminate "incomplete" episodes.) For bronchitis among children 0 to 7 years, episodes were constrained not to begin within four weeks of the end of either Period I or II to allow a full four weeks for a revisit to occur (see bronchitis, below).

If a visit for a study diagnosis occurred in the first two weeks of either period, then two weeks from the date of that visit were allowed to pass before a visit for the same diagnosis was used to begin an episode.

The "prescription rule" was implemented as follows for all oral medications:

1. If the prescription drug occurred on the same day as or followed a visit for one of the six study diagnoses or for one of the closely related diagnoses, it was given that diagnosis and thus included in the episode.
2. If it came one or two days before the initial visit in the episode, it was given the diagnosis of that initial visit and thus included in the episode.
3. If it occurred on the same day as, or followed, a visit for any other diagnosis, it was given that diagnosis and thus not included in the episode.

Laboratory tests and procedures were assigned to an episode on the basis of the diagnosis associated with that on a claim form. Because claims with two respiratory disease diagnoses were recoded to one (the more bacterial) diagnosis, the hierarchical rule was in effect for the services. If a laboratory test claim had no diagnosis, it was assigned by the same rules that governed prescription drugs (above). Strep and

throat cultures were included in all episodes of strep throat and pharyngitis regardless of diagnosis on the claim.

Once an episode was initiated by a visit, the data base was searched up to two days before that visit for all oral medications and for injections and other procedures for the diagnosis or a closely related diagnosis.

Rules for More Complex Episodes

Subsidiary rules were established by which changes could be made in episodes that appeared to be longer than two weeks in length and/or to have a change in diagnosis. This set applied only to the five study diagnoses other than otitis media.

1. Certain services that came after the end of a two week episode were included in the episode under special circumstances:

(a) If a visit for the same diagnosis as the initial visit in the episode occurred up to three days after the end of the episode (i.e., on Day 16, 17, or 18), and if no antibiotics were present on the day of the visit or up to two days afterwards, the visit was included in the previous episode. (If a visit for the same diagnosis as the initial visit in the episode occurred up to three days after the end of the episode but antibiotic medications were present on the same day as that visit or up to two days afterwards, it began a new episode, and the visit, antibiotics, and any other services on that day were included with the new episode.)

(b) Any other services that occurred on that same day as this "post-episode" visit (i.e., on Days 16, 17, or 18), and had the same diagnosis as the visit (i.e., the same diagnosis as the initial visit in the episode), were also assigned to the initial episode. (This could not include antibiotics or other services on a day with a visit and antibiotics, because a visit plus antibiotics on those days began a new episode.)

2. If the diagnoses in an episode "progressed" from less to more bacterial, a more complex way of creating the episode is invoked:

(a) The episode was continued for two weeks from the date of the "second" visit (i.e., the visit with the "more bacterial" diagnosis), and the usual rules for assigning services and medications to episodes applied. Services between Day 1 and the Day of the "second" (more bacterial) visit were still included.

(b) The episode was classified for analytic purposes as the more bacterial diagnosis (and quality of care was judged according to the "new" diagnosis).

(c) The physician type was changed (if necessary) to that corresponding to the visit of the more bacterial diagnosis.

(d) Only one such shift in diagnosis was allowed per episode.

(e) Services up to three days after the end of the extended episode (i.e., those pertaining to points 1(a) and 1(b), above) were not eligible to be part of an extended episode.

Followup Visits in Bronchitis

For bronchitis in children 0 to 7 years of age, a followup visit was defined as specified below:

1. A visit to any physician within four weeks (29 days) of the "initial visit for bronchitis" in any episode, where the followup visit has (a) any of the 11 study or closely related diagnoses, or (b) a diagnosis of well-child care (Y00.5). The initial visit for bronchitis can be (a) the first visit in an episode (i.e., an episode that actually began with bronchitis), or (b) a later visit in an episode that began with a less bacterial diagnosis (namely, influenza or acute URI).

2. A visit within four weeks (29 days) of the "initial visit for bronchitis" to the physician providing the initial visit, for any reason at all.

Followup visits in bronchitis for children 0 to 7 years of age also began new episodes in the following cases:

1. If a followup visit was beyond the usual 15-day (two-week) period (counting from the day of the bronchitis visit), and was for

strep throat, pharyngitis, influenza, or acute URI, it was used to start a new episode of that other diagnosis. (Because otitis media episodes were created separately, that diagnosis was irrelevant to this rule.)

2. If a followup visit beyond the 15-day period was for bronchitis and if antibiotic medications were given on the day of that visit or up to two days afterwards, that latter visit was also used to initiate a new episode of bronchitis.

Note that in both instances, a followup visit would still have been recorded as present for the preceding bronchitis episode.*

Otitis Media Rules

The first visit for otitis media for a given patient began an episode. Generally, all visits, laboratory tests, procedures, and medications in both oral and IM form that occurred on Day 1 through Day 15 and that were given for otitis media, for one of the other study diagnoses, or a closely related related diagnoses were included in the episode. Episodes of otitis media were constrained not to begin within two weeks of the beginning of either Period I or II, and not to begin within six weeks of the end of either Period I or II.

*Followup visits in bronchitis (0 to 7 years) determined the assignment of prescription drugs only when they occurred within the regular 15-day (two-week) period.

For otitis media, a followup visit was defined as follows:

(1) A visit to any physician within six weeks (or 43 days) of the initial visit in the episode for the following: (a) any otitis media code (i.e., 381.0-381.9), (b) any closely related diagnosis (which here included any of the five respiratory diagnosis as well), or (c) well-child care (Y00.5).

(2) A visit to the physician providing the initial visit for any reason at all.

Any otitis media followup visit between Day 16 and Day 43 that was associated with antibiotic medications on the day of the visit or up to two days afterward began a new episode of otitis media. (As with bronchitis, it was also counted as a followup visit in the previous episode.) Any otitis media visit occurring beyond the 43rd day of a previous otitis media episode began a new episode whether or not it was associated with antibiotics. Any of the rules in the previous sections that were not expressly overridden by the rules specific to otitis media were considered to apply to otitis media.

Antibiotic Combinations Variables

Certain combinations of the injectable and oral forms of antibiotics were considered separate items for the quality-of-care Profiles. Thus, once an episode was created according to the foregoing rules, the antibiotics assigned to it (alone or in combinations) were identified by separate variable numbers. The single antibiotics

included each of the five injectable (IM) antibiotics alone and each of the six oral antibiotics listed in (2) below, plus oral cephalosporin.

The possible combinations of interest included the following:

1. IM short-acting penicillin with the following oral antibiotics: penicillin, ampicillin, tetracycline, erythromycin, or lincomycin.
2. IM long-acting penicillin with the following oral antibiotics: penicillin, ampicillin, tetracycline, erythromycin, lincomycin, or sulfa.
3. IM ampicillin with oral penicillin or oral ampicillin.
4. IM lincomycin with the following oral antibiotics: penicillin, ampicillin, tetracycline, erythromycin, or lincomycin.
5. IM tetracycline with oral tetracycline.
6. Oral erythromycin with oral sulfa, and oral penicillin with oral sulfa.

All combinations other than those explicitly noted above, for example, pairs not otherwise mentioned, triplets, and so forth, were put in a residual combination category. This included, for example, all remaining combinations with tetracycline and lincomycin in either oral or injectable form, all other injectable antibiotic combinations without

any oral antibiotics, and all other oral antibiotic combinations without any injectable antibiotics.

No consideration was given as to which medication came first in an episode, and no counts of medications were made.

APPENDIX D

DIAGNOSIS-SPECIFIC DISTRIBUTIONS OF EPISODES BY LEVEL OF QUALITY OF CARE FOR SEPARATE CATEGORIES OF PROVIDERS IN PERIODS I AND II

The tables in this appendix give the number and percentage of episodes that were scored High, Acceptable, Minimal, and Unacceptable in both periods. For five of the respiratory conditions (strep throat, pharyngitis, bronchitis, influenza, and acute URI), the following categories of provider types are each given a separate table: certified group practices, noncertified group practices, and all group practices; certified MDs, nonoutlier (noncertified) MDs, outlier MDs, all noncertified MDs and all MDs; certified DOs, nonoutlier (noncertified) DOs, outlier DOs, all noncertified DOs, and all DOs. Tables D.1-D.13 are for Period I, D.14-D.26 for Period II.

For otitis media, the distributions based on both the original and the modified criteria (requiring and not requiring a followup visit, respectively) are given. The same provider categories are used. Tables D.27-D.29 are for Period I, D.30-D.32 for Period II.

The row labeled "Adequate" in each table (as explained in the text) is the subtotal of High, Acceptable, and Minimal care. Because percentages in each row were calculated separately, the subtotal Adequate may not in all instances be the same percentage as the sum of the individual percentages for High, Acceptable, and Minimal.

Table D.1
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 CERTIFIED GROUP PRACTICES, PERIOD I

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	35	49	25	20	7	18	19	70	128	58
Acceptable	17	24	73	57	14	37	0	0	18	8
Minimal	0	0	1	1	0	0	0	0	0	0
Adequate	52	73	99	77	21	55	19	70	146	66
Unacceptable	19	27	29	23	17	45	8	30	74	34
Total	71	100	128	100	38	100	27	100	220	100

250

Table D.2
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 NONCERTIFIED GROUP PRACTICES, PERIOD 1

Level of Care	Strep Throat Number Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
		Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	5	26	6	10	9	41	11	65	46
Acceptable	6	32	45	73	11	50	1	6	4
Minimal	0	0	0	0	0	0	0	0	0
Adequate	11	58	51	82	20	91	12	71	50
Unacceptable	8	42	11	18	2	9	5	29	61
Total	19	100	62	100	22	100	17	100	111

Table D.3
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 ALL GROUP PRACTICES, PERIOD I

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	40	44	31	16	16	27	30	68	174	53
Acceptable	23	26	118	62	25	42	1	2	22	7
Minimal	0	0	1	1	0	0	0	0	0	0
Adequate	63	70	150	79	41	68	31	70	196	59
Unacceptable	27	30	40	21	19	32	13	30	135	41
Total	90	100	190	100	60	100	44	100	331	100

Table D.4
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
CERTIFIED MDs, PERIOD I

Level of Care	Strep Throat Number	Throat Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	1	5	6	2	31	21	22	47	107	41
Acceptable	6	29	136	56	72	50	2	4	5	2
Minimal	0	0	2	1	0	0	1	2	1	1
Adequate	7	33	144	59	103	71	25	53	113	44
Unacceptable	14	67	100	41	42	29	22	47	145	56
Total	21	100	244	100	145	100	47	100	258	100

Table D.5
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
NONOUTLIER MDs, PERIOD I

Level of Care	Sore Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	1	3	13	2	70	31	35	39	165	40
Acceptable	16	52	294	55	99	43	0	0	15	4
Minimal	0	0	10	2	0	0	2	2	6	1
Adequate	17	55	317	59	169	74	37	42	186	45
Unacceptable	14	45	219	41	60	26	52	58	223	55
Total	31	100	536	100	229	100	89	100	409	100

Table D.6
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 OUTLIER MDs, PERIOD I

255

Level of Care	Strep Throat Number	Strep Throat Percent	Pharyngitis Number	Pharyngitis Percent	Bronchitis Number	Bronchitis Percent	Influenza Number	Influenza Percent	Acute URI Number	Acute URI Percent
High	0	0	3	1	18	14	7	10	40	22
Acceptable	8	12	70	17	27	20	0	0	0	0
Minimal	2	3	1	<1	2	2	2	3	2	1
Adequate	10	15	74	18	47	36	9	13	42	23
Unacceptable	58	85	330	82	85	64	59	87	139	77
Total	68	100	404	100	132	100	68	100	181	100

Table D.7
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 NONCERTIFIED MDs, PERIOD I

Level of Care	Strep Throat Number	Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	1	1	16	2	88	24	42	27	205	35
Acceptable	24	24	364	39	126	35	0	0	15	3
Minimal	2	2	11	1	2	1	4	3	8	1
Adequate	27	27	391	42	216	60	46	29	228	39
Unacceptable	72	73	549	58	145	40	111	71	362	61
Total	99	100	940	100	361	100	157	100	590	100

Table D.8
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
ALL MDS, PERIOD I

Level of Care	Strep Throat Number	Strep Throat Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	2	2	22	2	119	24	64	31	312	37
Acceptable	30	25	500	42	198	39	2	1	20	2
Minimal	2	2	13	1	2	<1	5	2	9	1
Adequate	34	28	535	45	319	63	71	35	341	40
Unacceptable	86	72	649	55	187	37	133	65	507	60
Total	120	100	1184	100	506	100	204	100	848	100

Table D.9
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 CERTIFIED DOs, PERIOD 1

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	0	0	0	0	5	12	2	40	15	56
Acceptable	2	40	47	31	14	35	0	0	0	0
Minimal	0	0	3	2	0	0	0	0	0	0
Adequate	2	40	50	33	19	48	2	40	15	56
Unacceptable	3	60	103	67	21	52	3	60	12	44
Total	5	100	153	100	40	100	5	100	27	100

Table D.10
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
NONOUTLIER DOS, PERIOD I

259

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	2	2	7	1	68	12	27	12	81	13
Acceptable	16	13	160	27	123	22	4	2	8	1
Minimal	0	0	3	1	7	1	2	1	6	1
Adequate	18	74	170	29	198	35	33	15	95	15
Unacceptable	109	86	420	71	367	65	191	85	549	85
Total	127	100	590	100	565	100	224	100	644	100

Table D.11
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
OUTLIER DOS, PERIOD 1

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	0	0	1	1	17	5	1	3	5	2
Acceptable	1	2	28	14	4	1	0	0	0	0
Minimal	0	0	1	1	0	0	0	0	0	0
Adequate	1	2	30	15	21	6	1	3	5	2
Unacceptable	48	98	168	85	323	94	37	97	216	98
Total	49	100	198	100	344	100	38	100	221	100

Table D.12
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
NONCERTIFIED DOS, PERIOD I

Level of Care	Strep Throat Number	Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	2	1	8	1	85	9	28	11	86	10
Acceptable	17	10	188	24	127	14	4	2	8	1
Minimal	0	0	4	1	7	1	2	1	6	1
Adequate	19	11	200	25	219	24	34	13	100	12
Unacceptable	157	89	588	75	690	76	228	87	765	88
Total	176	100	788	100	909	100	262	100	865	100

Table D.13
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 ALL DOS, PERIOD I

Level of Care	Strep Number	Throat Percent	Pharyngitis Number	Percent	Bronchitis Number	Percent	Influenza Number	Percent	Acute URI	
									Number	Percent
High	2	1	8	1	90	9	30	11	101	11
Acceptable	19	11	235	25	141	15	4	2	8	1
Minimal	0	0	7	1	7	1	2	1	6	1
Adequate	21	12	250	27	238	25	36	13	115	13
Unacceptable	160	88	691	73	711	75	231	87	777	87
Total	181	100	941	100	949	100	267	100	892	100

Table D.14
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 CERTIFIED GROUP PRACTICES, PERIOD II

263

Level of Care	Strep Throat Number	Strep Throat Percent	Pharyngitis Number	Pharyngitis Percent	Bronchitis Number	Bronchitis Percent	Influenza Number	Influenza Percent	Acute URI Number	Acute URI Percent
High	21	32	28	23	12	67	8	53	60	67
Acceptable	22	34	62	52	3	17	2	13	5	6
Minimal	0	0	3	2	1	6	0	0	0	0
Adequate	43	66	93	78	16	89	10	67	65	73
Unacceptable	22	34	27	22	2	11	5	33	24	27
Total	65	100	120	100	18	100	15	100	89	100

Table D.15
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
NONCERTIFIED GROUP PRACTICES, PERIOD II

Level of Care	Strep Number	Throat Percent	Pharyngitis Number	Percent	Bronchitis Number	Percent	Influenza Number	Percent	Acute URI Number	Percent
High	16	16	18	12	20	47	8	40	30	32
Acceptable	30	29	78	52	11	26	1	5	6	6
Minimal	1	1	1	1	2	5	2	10	4	4
Adequate	47	46	97	65	33	77	11	55	40	43
Unacceptable	55	54	53	35	10	23	9	45	53	57
Total	102	100	150	100	43	100	20	100	93	100

Table D.16
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 ALL GROUP PRACTICES, PERIOD II

Level of Care	Strep Throat Number	Strep Throat Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	37	22	46	17	32	52	16	46	90	49
Acceptable	52	31	140	52	14	23	3	9	11	6
Minimal	1	1	4	1	3	5	2	6	4	2
Adequate	90	54	190	70	49	80	21	60	105	58
Unacceptable	77	46	80	30	12	20	14	40	77	42
Total	167	100	270	100	61	100	35	100	182	100

Table D.17
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
CERTIFIED MDs, PERIOD II

Level of Care	Strep Throat Number	Strep Throat Percent	Pharyngitis Number	Pharyngitis Percent	Bronchitis Number	Bronchitis Percent	Influenza Number	Influenza Percent	Acute URI Number	Acute URI Percent
High	8	32	30	14	36	40	15	38	65	46
Acceptable	7	28	137	62	36	40	6	15	6	4
Minimal	0	0	1	<1	0	0	3	8	1	1
Adequate	15	60	168	76	72	80	24	60	72	51
Unacceptable	10	40	54	24	18	20	16	40	68	49
Total	25	100	222	100	90	100	40	100	140	100

Table D.18
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
NONOUTLIER MDs, PERIOD II

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	6	15	40	9	69	40	23	46	177	47
Acceptable	19	46	258	57	61	35	0	0	4	1
Minimal	0	0	11	2	3	2	1	2	3	1
Adequate	25	61	309	69	133	77	24	48	184	49
Unacceptable	16	39	141	31	40	23	26	52	190	51
Total	41	100	450	100	173	100	50	100	374	100

Table D.19
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
OUTLIER MDs, PERIOD II

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	0	0	1	0	51	57	21	30	86	52
Acceptable	8	50	122	37	11	12	1	1	3	2
Minimal	1	6	6	2	3	3	6	9	5	3
Adequate	9	56	129	39	65	72	28	41	94	57
Unacceptable	7	44	205	61	25	28	41	59	71	43
Total	16	100	334	100	90	100	69	100	165	100

268

Table D.20
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 NONCERTIFIED MDs, PERIOD II

Level of Care	Strep Throat Number	Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	6	11	41	5	120	46	44	37	263	49
Acceptable	27	47	380	48	72	27	1	1	7	1
Minimal	1	2	17	2	6	2	7	6	8	1
Adequate	34	60	438	56	198	75	52	44	278	52
Unacceptable	23	40	346	44	65	25	67	56	261	48
Total	57	100	784	100	263	100	119	100	539	100

Table D.21
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
ALL MDs, PERIOD II

Level of Care	Strep Throat Number	Pharyngitis Number	Bronchitis Number	Influenza		Acute URI	
				Percent	Number	Percent	Number
High	14	17	71	7	156	44	59
Acceptable	34	41	517	51	108	31	7
Minimal	1	1	18	2	6	2	10
Adequate	49	60	606	60	270	76	76
Unacceptable	33	40	400	40	83	24	83
Total	82	100	1006	100	353	100	159

Table D.22
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
CERTIFIED DOS, PERIOD II

Level of Care	Strep Number	Throat Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	0	0	5	5	29	52	3	43	5	42
Acceptable	1	6	16	17	17	30	0	0	1	8
Minimal	0	0	0	0	1	2	0	0	0	0
Adequate	1	6	21	22	47	84	3	43	6	50
Unacceptable	16	94	75	78	9	16	4	57	6	50
Total	17	100	96	100	56	100	7	100	12	100

Table D.23
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
NONOUTLIER DOS, PERIOD II

Level of Care	Strep Number	Throat Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	1	1	7	1	118	26	29	32	105	29
Acceptable	36	45	310	49	225	49	1	1	20	6
Minimal	2	2	10	2	13	3	3	3	4	1
Adequate	39	49	327	52	356	78	33	36	129	36
Unacceptable	41	51	300	48	99	22	59	64	227	64
Total	80	100	627	100	455	100	92	100	356	100

Table D.24
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
OUTLIER DOs, PERIOD II

Level of Care	Strep Throat Number	Throat Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	1	1	1	1	13	21	1	6	31	26
Acceptable	33	26	47	27	6	10	0	0	0	0
Minimal	1	1	10	6	3	5	0	0	6	5
Adequate	35	28	58	34	22	35	1	6	37	31
Unacceptable	92	72	115	66	41	65	15	94	83	69
Total	127	100	173	100	63	100	16	100	120	100

Table D.25
NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
NONCERTIFIED DOs, PERIOD II

Level of Care	Strep Throat		Pharyngitis		Bronchitis		Influenza		Acute URI	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	2	1	8	1	131	25	30	28	136	29
Acceptable	69	33	357	45	231	45	1	1	20	4
Minimal	3	1	20	2	16	3	3	3	10	2
Adequate	74	36	385	48	378	73	34	31	166	35
Unacceptable	133	64	415	52	140	27	74	69	310	65
Total	207	100	800	100	518	100	108	100	476	100

Table D.26
 NUMBER AND PERCENT OF EPISODES WITH HIGH TO UNACCEPTABLE CARE:
 ALL DOS, PERIOD II

Level of Care	Strep Throat Number	Throat Percent	Pharyngitis		Bronchitis		Influenza		Acute URI	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
High	2	1	13	1	160	28	33	29	141	29
Acceptable	70	31	373	42	248	43	1	1	21	4
Minimal	3	1	20	2	17	3	3	3	10	2
Adequate	75	33	406	45	425	74	37	32	172	35
Unacceptable	149	67	490	55	149	26	78	68	316	65
Total	224	100	896	100	574	100	115	100	488	100

Table D.27

NUMBER AND PERCENT OF EPISODES OF OTITIS MEDIA WITH HIGH TO UNACCEPTABLE CARE:
GROUP PRACTICES, PERIOD I

Otitis Media Scores ^a	Certified Groups			Noncertified Groups			All Groups		
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
<u>Original</u>									
High	5	6	1	5	5	6	6	3	6
Acceptable	3	4	0	0	0	0	3	2	3
Minimal	2	3	0	0	0	0	2	1	2
Adequate	10	13	1	5	11	11	11	11	11
Inadequate	67	87	18	95	85	89	85	89	89
Total	77	100	19	100	96	100	96	100	100
<u>Revised</u>									
High	5	6	1	5	5	6	6	3	6
Acceptable	3	4	0	0	0	0	3	3	3
Minimal	9	12	5	26	26	26	14	15	15
Adequate	17	32	6	32	23	24	23	24	24
Inadequate	60	68	13	68	73	76	73	76	76
Total	77	100	19	100	96	100	96	100	100

^a Original means that minimal required both a followup visit and appropriate antibiotics; revised means that minimal required only appropriate antibiotics.

Table D.28
NUMBER AND PERCENT OF EPISODES OF OTITIS MEDIA WITH HIGH TO UNACCEPTABLE CARE:
MDs, PERIOD 1

Otitis Media Scores ^a	Certified MDs		Nonoutlier		Noncertified MDs		All MDs	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
<u>Original</u>								
High	10	15	4	6	0	0	4	8
Acceptable	2	3	2	3	1	3	5	3
Minimal	0	0	2	3	1	3	3	2
Adequate	12	18	8	11	2	6	10	13
Inadequate	56	82	63	89	34	94	97	87
Total	68	100	71	100	36	100	107	100
<u>Revised</u>								
High	10	15	4	6	0	0	4	14
Acceptable	2	3	2	3	1	3	3	5
Minimal	14	21	23	32	9	25	32	30
Adequate	26	38	29	41	10	28	39	36
Inadequate	42	62	42	59	26	72	68	64
Total	68	100	71	100	36	100	107	100

^aOriginal means that minimal required both a followup visit and appropriate antibiotics; revised means that minimal required only appropriate antibiotics.

Table D.29

NUMBER AND PERCENT OF EPISODES OF OTITIS MEDIA WITH HIGH TO UNACCEPTABLE CARE:
DOs, PERIOD I

Otitis Media Scores ^a	Certified DOs		Nonoutlier		Noncertified DOs		All DOs	
	Number	Percent	Number	Percent	Outlier Number	Percent	Total Number	Percent
<u>Original</u>								
High	1	3	3	3	0	0	3	3
Acceptable	1	3	1	1	0	0	1	4
Minimal	0	0	0	0	0	0	0	3
Adequate	2	7	4	4	0	0	0	1
Inadequate	27	93	88	96	19	100	107	6
Total	29	100	92	100	19	100	111	100
<u>Revised</u>								
High	1	3	3	3	0	0	3	3
Acceptable	1	3	1	1	0	0	1	4
Minimal	10	34	10	11	0	0	10	1
Adequate	12	41	14	15	0	0	14	20
Inadequate	17	59	78	85	19	100	97	26
Total	29	100	92	100	19	100	111	100

^aOriginal means that minimal required both a followup visit and appropriate antibiotics; revised means that minimal required only appropriate antibiotics.

Table D.30
NUMBER AND PERCENT OF EPISODES OF OTITIS MEDIA WITH HIGH TO UNACCEPTABLE CARE:
GROUP PRACTICES, PERIOD II

Otitis Media Scores ^a	Certified Groups		Noncertified Groups		All Groups	
	Number	Percent	Number	Percent	Number	Percent
<u>Original</u>						
High	4	18	1	5	5	12
Acceptable	3	14	2	11	5	12
Minimal	2	9	0	0	2	5
Adequate	9	41	3	16	12	29
Inadequate	13	59	16	84	29	71
Total	22	100	19	100	41	100
<u>Revised</u>						
High	4	18	1	5	5	12
Acceptable	3	14	2	11	5	12
Minimal	5	23	2	11	7	17
Adequate	12	55	5	26	17	41
Inadequate	10	45	14	74	24	59
Total	22	100	19	100	41	100

^aOriginal means that minimal required both a followup visit and appropriate antibiotics; revised means that minimal required only appropriate antibiotics.

Table D.31

NUMBER AND PERCENT OF EPISODES OF OTITIS MEDIA WITH HIGH TO UNACCEPTABLE CARE:
MDS, PERIOD II

Otitis Media Scores ^a	Certified MDs Number	Certified MDs Percent	Nonoutlier			Noncertified MDs			All MDs		
			Number	Percent	Number	Percent	Number	Percent	Number	Percent	
<u>Original</u>											
High	5	14	3	14	0	0	3	9	8	12	
Acceptable	2	6	2	10	0	0	2	6	4	6	
Minimal	0	0	0	0	0	0	0	0	0	0	
Adequate	7	20	5	24	0	0	5	15	12	18	
Inadequate	28	80	16	76	12	100	28	85	56	82	
Total	35	100	21	100	12	100	33	100	68	100	
<u>Revised</u>											
High	5	14	3	14	0	0	3	9	8	12	
Acceptable	2	6	2	10	0	0	2	6	4	6	
Minimal	7	20	8	38	8	67	16	48	23	34	
Adequate	14	40	13	62	8	67	21	64	35	51	
Inadequate	21	60	8	38	4	33	12	36	33	49	
Total	35	100	21	100	12	100	33	100	68	100	

^aOriginal means that minimal required both a followup visit and appropriate antibiotics; revised means that minimal required only appropriate antibiotics.

Table D.32
NUMBER AND PERCENT OF EPISODES OF OTITIS MEDIA WITH HIGH TO UNACCEPTABLE CARE:
DOs, PERIOD II

Otitis Media Scores ^a	Certified DOs		Nonoutlier		Noncertified DOs		All DOs	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
<u>Original</u>								
High	0	0	0	0	0	0	0	0
Acceptable	0	0	4	7	1	8	5	7
Minimal	0	0	0	0	1	8	1	1
Adequate	0	0	4	7	2	15	6	6
Inadequate	4	100	55	93	11	85	66	92
Total	4	100	59	100	13	100	72	100
<u>Revised</u>								
High	0	0	0	0	0	0	0	0
Acceptable	0	0	4	7	1	8	5	7
Minimal	2	50	30	51	4	31	34	47
Adequate	2	50	34	58	5	38	39	54
Inadequate	2	50	25	42	8	62	33	46
Total	4	100	59	100	13	100	72	100
281								

a. Original means that minimal required both a followup visit and appropriate antibiotics; revised means that minimal required only appropriate antibiotics.

AD-A103 341 RAND CORP SANTA MONICA CA
QUALITY OF CARE IN EPISODES OF COMMON RESPIRATORY INFECTIONS IN--ETC(U)
OCT 80 K N LOHR

F/G 6/5

UNCLASSIFIED RAND/P-6570

NL

4 or 4
AD-A
103-341

END
DATE
10-81
DTIC

APPENDIX E

USE OF THROAT CULTURES AND APPROPRIATE ANTIBIOTICS IN
SORE THROAT EPISODES FOR SEPARATE CATEGORIES OF PROVIDERS IN
PERIODS I AND II

Appendix E, Table E.1

NUMBER AND PERCENT OF SORE THROAT EPISODES WITH THROAT CULTURES
AND/OR APPROPRIATE ANTIBIOTICS, BY TYPE OF PROVIDER: PERIOD 1

Type of Provider	Throat Culture and Appropriate Antibiotics		Throat Culture but not Appropriate Antibiotics		Appropriate Antibiotics but no Throat Culture		Neither Throat Culture Nor Appropriate Antibiotics		Total	
	N	%	N	%	N	%	N	%	N	%
Groups	60	21	42	15	123	44	55	20	280	100
MDs	26	2	16	1	514	39	748	57	1304	99
DOS	1	>1	8	1	245	22	868	77	1122	100
Certified Groups	51	26	38	19	73	37	37	19	199	101
Noncertified Groups	9	11	4	5	50	62	18	22	81	100
Certified MDs	6	2	4	2	138	52	117	44	265	100
Noncertified MDs	20	2	12	1	376	36	631	61	1039	100
Certified DOS	0	-	2	1	46	29	110	70	158	100
Noncertified DOS	1	>1	6	1	199	21	758	79	964	101
Nonoutlier MDs	17	3	12	2	303	53	235	41	567	99
Outlier MDs	3	1	0	-	73	15	396	84	472	100
Nonoutlier DOS	1	>1	4	1	169	24	543	76	717	101
Outlier DOS	0	-	2	1	30	12	215	87	247	100

^aPercentages may not sum to 100 because of rounding.

Appendix E, Table E.2
NUMBER AND PERCENT OF SORE THROAT EPISODES WITH THROAT CULTURES
AND/OR APPROPRIATE ANTIBIOTICS, BY TYPE OF PROVIDER: PERIOD II

Type of Provider	Throat Culture		Throat Culture and Appropriate Antibiotics		Appropriate Antibiotics but no Throat Culture		Neither Throat Culture Nor Appropriate Antibiotics		Total	
	N	%	N	%	N	%	N	%	N	%
			Antibiotics	Antibiotics	Throat Culture	Antibiotics	Throat Culture	Antibiotics	Throat Culture	Antibiotics
Groups	80	18	69	16	172	39	116	27	437	100
MDs	91	8	43	4	538	49	416	38	1088	99
DOs	31	3	63	6	424	38	602	54	1120	101
Certified Groups	46	25	33	18	69	37	37	20	185	100
Noncertified Groups	34	13	36	14	103	41	79	31	252	99
Certified MDs	40	16	16	6	136	55	55	22	247	99
Noncertified MDs	51	6	27	3	402	48	361	43	841	100
Certified DOs	5	4	3	3	15	13	90	80	113	100
Noncertified DOs	26	3	60	6	409	41	512	51	1007	101
Nonoutlier MDs	50	10	27	5	268	55	146	30	491	100
Outlier MDs	1	>1	0	-	134	38	215	61	350	99
Nonoutlier DOs	26	4	59	8	322	46	300	42	707	100
Outlier DOs	0	-	1	>1	87	29	212	71	300	100

^aPercentages may not sum to 100 because of rounding.

REFERENCES

Ainsworth, T. H., Jr. Quality Assurance in Long Term Care.

Germantown, MD: Aspen Systems Corporation, 1977.

Avery, A. D., Lelah, T., Solomon, N. E., et al. Quality of Medical Care Assessment Using Outcome Measures: Eight Disease-Specific Applications. R-2021/2-HEW. Santa Monica, CA, The Rand Corporation, 1976.

Brook, R. H. Quality of Care Assessment: A Comparison of Five Methods of Peer Review. DHEW publication no. HRA 74-2100. Washington, DC; Government Printing Office, 1973. (Rockville, MD; National Center for Health Services Research and Development, Department of Health, Education, and Welfare, 1973).

Brook, R. H., and Davies-Avery, A. Mechanisms for Assuring Quality of U.S. Medical Care Services: Past, Present and Future. R-1939-HEW. Santa Monica, CA, The Rand Corporation, 1977.

Brook, R. H., Davies-Avery, A., Greenfield, S., et al. Assessing the Quality of Medical Care Using Outcome Methods: An Overview of the Methods. Medical Care 15:1-65, supplement, 1977.

Brook, R. H., Lohr, K. N., Goldberg, G. A., et al. Conceptualization and Measurement of Physiologic Health for Adults. R-2262-HHS

(multiple volumes of a Health Insurance Experiment Series). Santa Monica, CA, The Rand Corporation, 1980-.

Brook, R. H., and Williams, K. N. Quality of Health Care for the Disadvantaged. Journal of Community Health 1:132-156, 1975.

Brook, R. H., and Williams, K. N. Evaluation of the New Mexico Peer Review System. Medical Care 14:1-122, supplement, (December) 1976.

Brook, R. H., Williams, K. N., and Rolph, J. E. Controlling the Use And Cost of Medical Services: The New Mexico Experimental Medical Care Review Organization: A Four Year Study. Medical Care 16:1-76, supplement, (September) 1978.

CBO. The Effect of PSROs on Health Care Costs: Current Findings and Future Evaluations. Washington, DC; Congressional Budget Office, U.S. Congress, 1979.

Christoffel, T., and Loewenthal, M. Evaluating the Quality of Ambulatory Health Care: A Review of Emerging Methods. Medical Care 16:887-897, 1977.

Covell, R. The Impact of Regulation on Health Care Quality. In Levin, A., ed. Regulating Health Care. The Struggle for Control. Proceedings of the Academy of Political Science, Vol. 33, No. 4. New York: The Academy, 1980.

Donabedian, A. Needed Research in the Assessment and Monitoring of the Quality of Medical Care. DHEW publication no. PHS 78-3219. Hyattsville, MD; National Center for Health Services Research, 1978.

Dutton, D. B., and Silber, R. S. Children's Health Outcomes in Six Different Ambulatory Care Delivery Systems. Medical Care 18:693-714, 1980.

Egdahl, R. H., and Gertman, P. M., eds. Quality Assurance in Health Care. Germantown, MD: Aspen Systems Corporation, 1976.

Egdahl, R. H., and Gertman, P. M., eds. Technology and the Quality of Health Care. Germantown, MD: Aspen Systems Corporation, 1978.

Farrington, J. F., Felch, W. C., and Hare, R. L. Quality Assessment and Quality Assurance. The Performance-Review Alternative. New England Journal of Medicine 303:154-156, 1980.

Galblum, T. W., compiler. Research and Demonstration in Health Care Financing, 1978-1979. HCFA publication no. 03044. Baltimore, MD; Department of Health and Human Resources, Health Care Financing Administration, no date.

Gertman, P. M., Monheit, A. C., Anderson, J. J., et al. Utilization Review in the United States: Results from a 1976-1977 National Survey of Hospitals. Medical Care 17:1-147, supplement, (August) 1979.

Gold, M. Relative Intensity of Hospital-Based Primary Care.

Unpublished Sc.D. thesis, Harvard School of Public Health, 1979.

Goldberg, G. A., and Jolly, D. G. Quality of Care Provided by Physician's Extenders in Air Force Primary Medical Clinics.

R-2436-AF. Santa Monica, CA, The Rand Corporation, 1980.

Goran, M. J. The Evaluation of the PSRO Hospital Review System.

Medical Care 17:1-47, supplement, (May) 1979.

Goran, M. J., Roberts, J. S., Kellogg, M. A., et al. The PSRO Review System. Medical Care 13:1-33, supplement, 1975.

Greene, R. Assuring Quality of Medical Care. Cambridge, MA: Ballinger Publishing Company, 1976.

Greenfield, S., Kaplan, S. H., Goldberg, G. A., et al. Physician Preference for Criteria Mapping in Medical Care Evaluation. Journal of Family Practice 6:1079-1086, 1978.

Hare, R. L., and Barnoon, S. Medical Care Appraisal and Quality Assurance in the Office Practice of Internal Medicine. San Francisco, CA: American Society of Internal Medicine, 1973.

HCFA. Professional Standards Review Organization. 1979 Program

Evaluation. HCFA publication no. 03041. Baltimore, MD: Department of Health and Human Services, Health Care Financing Administration, 1980.

Hulka, B. S., Romm, F. J., Parkerson, G. R., Jr., et al. Peer Review and Ambulatory Care: Use of Explicit Criteria and Implicit Judgments. Medical Care 17:1-73, supplement, (March) 1979.

IOM. Advancing the Quality of Health Care: Key Issues and Fundamental Principles. Washington, DC; National Academy of Sciences, Institute of Medicine, 1974.

IOM. Assessing Quality in Health Care: An Evaluation. Washington, DC; National Academy of Sciences, Institute of Medicine, 1976.

Kane, R. A., Kane, R. L., Kleffel, D., et al. The PSRO and the Nursing Home. R-2459-HCFA. Vol. I: An Assessment of PSRO Long-Term Care Review. Vol. II: Ten Demonstration Projects in PSRO Long-Term Care Review. Santa Monica, CA, The Rand Corporation, 1979.

Kane, R. L., and Kane, R. A. Long-Term Care: Can Our Society Meet the Needs of Its Elderly? Annual Review of Public Health 1:227-253, 1980.

Kane, R. L. Gardner, J. Wright, D. D., et al. A Method for Assessing the Outcome of Acute Primary Care. Journal of Family Practice 4:1119-1124, 1977.

Kane, R. L., Gardner, J., Wright, D. D. et al. Differences in the Outcomes of Acute Episodes of Care Provided by Various Types of Family Practitioners. Journal of Family Practice 6:133-138, 1978.

Kane, R. L., Olsen, D. M., Castle, D. H. Medex and Their Physician Predictors. Journal of the American Medical Association 236:2509-2512, 1976a.

Kane, R. L., Woolley, F. R., Gardner, J., et al. Measuring Outcomes of Care in an Ambulatory Primary Care Population. Journal of Community Health 1:233-240, 1976b.

Kessner, D. M. Quality Assessment and Assurance: Early Signs of Cognitive Dissonance. New England Journal of Medicine 298:381-386, 1978.

Lasdon, G. S., and Sigmann, P. Evaluating Cost-Effectiveness Using Episodes of Care. Medical Care 15:260, 1977.

Lewis, C. E. State of the Art of Quality Assessment. Medical Care 12:799-806, 1974.

Loehr, K. N., Brook, R. E., and Kaufman, M. A. Quality of Care in the New Mexico Medicaid Program. The Effect of the New Mexico Experimental Medical Care Review Organization on the Use of Antibiotics for Common Infectious Diseases. Medical Care 18:1-129, supplement, (January) 1980.

McAuliffe, W. E. Measuring Quality of Medical Care: Process Versus Outcome. Milbank Memorial Fund Quarterly/Health and Society 57:118-152, 1979.

Miller, M. C., and Knapp, R. G. Evaluating Quality of Care. Analytic Procedures--Monitoring Techniques. Germantown, MD: Aspen Systems Corporation, 1979.

Moscovice, I. A Method for Analyzing Resource Use in an Ambulatory Care Setting. Medical Care 15:1024, 1977.

Mushlin, A. E., and Appel, F. A. Testing an Outcome-Based Quality Assurance Strategy in Primary Care. Medical Care 18:1-100, supplement, (May) 1980.

Newhouse, J. P. A Design for a Health Insurance Experiment. Inquiry 11:5-27, 1974.

OPEL. PSRO: An Evaluation of the Professional Standards Review Organization: Executive Summary. Draft of DHEW publication no. OPEL#77-12, Vol. 1. Washington, DC: DHEW/HSA, Office of Planning, Evaluation, and Legislation, 1977.

Osborne, C. E., and Thompson, H. E. Criteria for Evaluation of Ambulatory Child Health by Chart Audit: Development and Testing of a Methodology. Pediatrics 56 (Pt. 2, Supplement):665-692, 1975.

OTA. The Implications of Cost-Effectiveness Analysis of Medical Technology. OTA-H-126 . Washington, D.C.: Congress of the United States, Office of Technology Assessment, 1980.

Paris, M., McNamara, J., and Schwartz, M. Monitoring Ambulatory Care: Impact of a Surveillance Program on Clinical Practice Patterns in New York City. American Journal of Public Health 70:783-788, 1980.

Payne, B. C., Lyons, T. F., Dwarshius, L., et al. The Quality of Medical Care: Evaluation and Improvement. Chicago: Hospital Research and Educational Trust, 1976.

Ray, W. A., Federspiel, C. F., and Schaffner, W. Prescribing of Chloramphenicol in Ambulatory Practice. Annals of Internal Medicine 84:266-270, 1976.

Ray, W. A., Federspiel, C. F., and Schaffner, W. The Mal-Prescribing of Liquid Tetracycline Preparations. American Journal of Public Health 67:762-763, 1977a.

Ray, W. A., Federspiel, C. F., and Schaffner, W. Prescribing of Tetracycline to Children Less than 8 Years Old. Journal of the American Medical Association 237:2069-2074, 1977b.

Riedel, R. L., and Riedel, D. C. Practice and Performance: An Assessment of Ambulatory Care. Ann Arbor, MI: Health Administration Press, 1979.

Romm, F. J., and Hulka, B. S. Developing Criteria for Quality-of-Care Assessment: Effect of the Delphi Technique. Health Services Research 14:309-312, 1979.

Roos, L. L., Jr., Roos, N. P., Henteleff, P. D. Assessing the Impact of Tonsillectomies. Medical Care 16:64-81, 1978.

Roos, L. L., Jr., Nicol, J. P., Johnson, C. F., et al. Using Administrative Data Banks for Research and Evaluation. Evaluation Quarterly 3:236-255, 1979.

Roos, N. P., Henteleff, P. D., Roos, L. L., Jr. A New Audit Procedure Applied to an Old Question: Is the Frequency of T & A Justified? Medical Care 15:1, 1977a.

Roos, N. P., Roos, L. L., Jr., Henteleff, P. D. Elective Surgical Rates--Do High Rates Mean Lower Standards? New England Journal of Medicine 297:360-365, 1977b.

Sanazaro, P. J. Quality Assessment and Quality Assurance in Medical Care. Annual Review of Public Health 1:37-68, 1980.

Schaffner, W., Ray, W. A., and Federspiel, C. F. Surveillance of Antibiotic Prescribing in Office Practice. Annals of Internal Medicine 89:796-799, 1978.

Solon, J. A., Feeney, S. J., Jones, S. H., et al. Delineating Episodes of Medical Care. American Journal of Public Health 57:401-408, 1967.

Solon, J. A., Rigg, R. D., Jones, S. H., et al. Episodes of Medical Care: Nursing Students' Use of Medical Services. American Journal of Public Health 59:936-946, 1969.

Sullivan, R. J., Estes, E. H., Stopford, W., et al. Adherence to Explicit Strategies for Common Medical Problems. Medical Care 18:388-399, 1980.

Toman, R. R., White, N. H., Giebink, G. A. Ambulatory Care Quality Assurance Project. Bibliography and Selected Abstracts. DHEW publication no. HSA#77-300, Vol. 3 Washington DC: U.S. Government Printing Office, 1976.

Tonkin, R. S. The REACH Centre: Its History and Work. III. A Study of the Quality of Pediatric Care. Canadian Journal of Public Health 70:405-414, 1979.

Ware, J. E., Jr., Brook, R. H., Davies-Avery, A., et al. Conceptualization and Measurement of Health for Adults in the Health Insurance Study. Vol. I. R-1987/1-HEW. Santa Monica, CA, The Rand Corporation, 1979.

Williams, K. N. Implementation of Professional Standards Review Organizations: A Theoretical Discussion. P-5757. Santa Monica, CA, The Rand Corporation, 1976.

Williams, K. N., and Brook, R. H. Quality of Measurement and Assurance. A Review of the Recent Literature. Health and Medical Care Services Review 1:1-15, May/June, 1978.

Williamson, J. W. Improving Medical Practice and Health Care. A Bibliographic Guide to Information Management in Quality Assurance and Continuing Education. Cambridge, MA: Ballinger Publishing Company, 1977.

Williamson, J. W. Assessing and Improving Outcomes in Health Care: The Health Accounting Approach to Quality Assurance. Cambridge, MA: Ballinger Publishing Company, 1978.

N
DAI